

More money rows for academics

In what seems certain to be another year of introspection, British universities will have to take a more open view on the future pattern of student support.

THOSE who have just swum the English Channel are rarely pleased to be told that they must turn round and swim back again. That is how the British university system will consider it is being dealt with, in the political year now beginning. After the trauma of the Education Reform Bill, when universities lost the right to offer teachers indefinite tenure but successfully won concessions in the cause of academic freedom, they might have expected a breathing-spell. Instead, they will find that their paymaster, the British government, is looking for a continuation of what it calls reform by administrative rather than legislative means. Having reshaped the mechanisms by which public subventions of the system are disbursed (the University Grants Committee will indeed be replaced by the Universities Funding Council next April, for example), the government is bent on redistributing the amounts of money available under different headings. But, on this occasion, the university system may better the situation in which it finds itself if it plays its cards well, and may even win back some of the freedom lost in recent years.

There are three sources of public support for British universities. First, there is a general subvention of £1,860 million (this year) channelled through the University Grants Committee, of which a third is reckoned to cover the infrastructure of research. Second, there is the so-called Science Budget, £670 this year, most of which is channelled through the research councils as research grants, services on behalf of university research in general and stipends for graduate students; this chunk of the budget is controlled by the Department of Education and Science (DES) on the advice of the Advisory Board for the Research Councils (ABRC). Third, there is spending in respect of university tuition fees and maintenance grants for undergraduates channelled through local governments, which pay fees for and maintenance grants to students in approved higher education courses according to rules laid down by DES, which reimburses local authorities.

In the university sector, tuition fees account for roughly £120 million a year, and maintenance grants to students for perhaps £250 million a year. The total is far from insignificant. Now that the legislative dust has settled, it is natural that all interested parties should be asking whether the same money spent differently might yield a better system.

Inclinations

On one point, the government's inclinations are already plain. Mr Kenneth Baker, the Secretary of State for Education and Science, has been hinting that he would welcome a transfer of funds from institutional to research support, or from UGC to the research councils. The effect would be to end what the British call the 'dual-support' system, in which universities finance the continuing capability of departments for research and the research councils support special projects. So deeply ingrained is the system that the idea that it might be abandoned was first canvassed only six years ago in a joint UGC-ABRC study, although UGC's more recent scheme to skew general university support towards departments with good records in research means that some are now distinctly more equal than others.

But there are two clear dangers in the wholehearted pursuit of

this policy. First, universities whose general support covers only their teaching will be less able to encourage newly promising departments to embark on adventures in research. Second, a transfer of funds will not mean that the research councils will be able to support more research projects; they will merely become enmeshed in haggles about the level of overhead payments.

The other budget point on which the government's attention is riveted is its mechanism for supporting students. Britain stands out among otherwise comparable countries in its general undertaking to feed and house students in higher education. That it does so in a niggardly fashion, with maintenance grants too small to keep body and soul together even when they are not reduced below the maximum (often to zero) by a calculation of a student's parents' income, does not prevent it claiming to be generous. But the practice has come to seem increasingly anomalous as British prosperity has been restored. Although British students are less able than elsewhere to help support themselves with part-time earnings, it must surely be sensible to ask whether the university system could be better managed if the relatively large sum of £250 million, or even a large part of it, were differently spent.

This is where there are opportunities. DES is not well-placed to engineer a substantial change of policy without assistance. So much is clear from the drubbing Sir Keith Joseph, Baker's predecessor, received from his own supporters when he tried to introduce a scheme that would have required well-to-do parents not only to maintain their offspring in higher education but also to pay nominal tuition fees. There is little sign that the government's own supporters have been weaned from their dependence on student maintenance and nominal university fees. Nor, for that matter, have academics, who regard the continuation of the present system as a necessary assurance, in a hard world, that there will be students to teach.

In reality, it is now probably safe to calculate differently. The British national interest and that of the university system coincide at one point, that there should be an increase in the proportion of young people in higher education so as the better to simulate the performance of the United States and Japan. The British government would more readily agree if it were not, under present rules, required to pay maintenance awards to students. So is there some way in which this small part of the whole cake can be resliced so as to give the universities more freedom and flexibility while increasing the chance that young people from poor families, already under-represented in higher education, will be more ready to follow higher education courses? One obvious component is to use part of the money for schemes for scholarship assistance administered by universities, not by the government or local authorities. Partly so as to meet the cost, but also as a matter of equity, there should also be a system of student loans. Hitherto, the opposition has been grounded on the view that loans are inequitable, and a disincentive for poor students. But circumstances have changed, and British universities have no way of telling how much academic business they are now losing because there is no realistic way in which students can win a professional education for themselves except through the mechanisms of the welfare state. □



Generations of chips

A decade of upheaval is about to be followed by another, and then another. . . .

JUST when it becomes clear that the industrialized world's working life has been transformed by the ready availability of personal computers, it begins to be apparent that another wave of a no doubt recurring revolution is about to engulf us. So much is signalled by the steady dribble of news of technical innovation in the computer industry stemming largely from the United States. There is a sense in which what has happened so far amounts merely to a demonstration that individual desks can be equipped with the computing power for which large organizations were eager to pay 100 to 1,000 times as much as recently as the 1970s. Now, there is a prospect that next decade's desks will carry such great computing power that, for practical purposes and for most users, it will be essentially infinite.

Nobody should be surprised. A little reflection will show that the technical basis of the past decade's transformation is only narrow. The engines of change so far have been two families of microprocessors manufactured in the United States by the Intel Corporation and Motorola Inc. Those with personal computers on their desks might usefully reflect on the intellectual effort embodied in these devices, which hang on simulations on a single piece of silicon of the electronic networks which, in the 1970s, were known as mainframe computers; mass-production may have made these computing elements cheap, but one cannot but marvel that so much cleverness has been made affordable. That such a device will, of its own accord, interpret the verbal jumble 'MOV AX, *string*' as the instruction to load its accumulator with the address in memory of a previously defined variable called *string* makes available to ordinary mortals the experience accumulated over a quarter of a century by professional programmers.

Yet ordinary users are not required to be familiar with these intricacies. The pace of the past decade's change has been sustained by the commercial development of software programmes for performing large collections of cognate routine tasks without knowing much about the working of the machinery, much as people can drive vehicles successfully without knowing how carburettors work.

The successes are plain. The numbers of personal computer systems sold each year are of the order of one per cent of the working population of the industrialized world. Most are versions of the personal computers developed by International Business Machines Inc., but Apple Computer Inc. let it be known two weeks ago that it sold 900,000 copies of its distinctive Macintosh II system during its last financial year. It is also clear that, in the drive for general accessibility by hardware and software manufacturers, much of the technical potential even of existing hardware has been neglected; the fraction of users of the AT version of IBM's personal computer exploiting the full capacity of its Intel 80246 chip, for example, must be very small, while only now is the software for the more sophisticated 80386 chip, which handles 32-bit words, becoming available.

Meanwhile, the software that made the unimagined possible a decade ago comes to seem much less sophisticated than it might be. The routine functioning of most existing personal computers is guided by a version of IBM's operating system, originally PC-DOS, then MS-DOS. But opinion is now shifting to the view that the UNIX operating system developed by AT&T has advantages in handling large quantities of information; AT&T has joined with the US company called Sun Computers to develop a version of UNIX for general use, while its competitors (including IBM) are jointly working on an alternative. (Common sense would persuade the two camps to bury their rivalry, but that may not be the outcome.) Yet the software people and their users are already asking whether information need necessarily be organized linearly, as files.

Even before these issues have been settled, the personal computer business will be overtaken by new hardware developments. Work-stations differ from most personal computers in that a single microprocessor can orchestrate a much larger array of rapid-access memory (RAM, which stands for random-access memory). They are certain to capture a substantial part of the present market for graphics use and are likely to be at once affordable and, in some ways, simpler to use. Desk-top parallel computer systems are also on the way; witness the plan announced last week by Mr Stephen E. Jobs, one of Apple Computer's founders, to manufacture such a system, initially for educational users. It may be the best part of a decade before there will have been the intellectual effort required fully to exploit these developments, by which time the hardware manufacturers will no doubt be working on the chips suggested by recent developments in artificial intelligence and neurophysiology. Even as the new wave of revolution threatens, the beginnings of its successor can be discerned. □

Tobacco for sale

Reorganization in the tobacco industry suggests that business is coming to an end.

THE tobacco industry has gone a long way, in the past week, towards the recognition of its own eventual demise. That is the implication of the proposed reorganization of two of the largest tobacco manufacturers in the United States. For several years now, the weed-dispensers have been busily diversifying their interests. Less obviously damaging foodstuffs have been natural choices; the retail outlets are often the same, and there is a good chance that the same sales forces can win orders for both kinds of products which is what stock markets call synergy. Now, two manufacturers have devised schemes that will salvage something for shareholders from the eventual collapse of tobacco sales. Philip Morris, which makes Marlboro cigarettes, is offering to buy for a cool \$11,000 million the company called Kraft, best known for its rubber-like sliced cheese. And the management of RJR Nabisco, itself the result of an earlier merger between a tobacco company (R.J. Reynolds) and a food company (Nabisco), is offering to buy out other shareholders at a cost of no less than \$17,000 million.

Although the objectives of the two schemes are very different, the underlying assumptions are the same. Cigarette sales in the United States are now declining by roughly 2 per cent a year. Moreover, the cost of manufacturing cigarettes is only about a third of the revenues the manufacturers collect. Inevitably, there is also a substantial profit, nearly \$2,000 million a year in RJR Nabisco's case, which can be used to finance the purchase of some other business through the now-familiar financial instruments called 'junk bonds'. That is how Philip Morris would pay for Kraft, and how in the short run the management of RJR Nabisco would buy out the other shareholders (but it might then sell Nabisco, keeping the cigarette money for itself for the duration of this sordid business).

It will be for the stock markets to decide whether the bidders have offered enough for what they seek to buy. A more awkward question underlying the calculations is whether the tobacco market will hold up for long enough to enable the bidders to service the junk bonds they will issue. Part of the inspiration of both deals, and one of the reasons why the financial calculations are especially chancy, is that the climate for the tobacco companies has significantly worsened in the past year. The first product liability suit against a tobacco manufacturer has been lost in New Jersey (but the jury verdict is being appealed against). The spread of municipal regulations prohibiting tobacco smoke in public and even, sometimes, private places is a more telling reminder to shareholders that their assets are wasting assets. They will do less well than if they had sold some years ago, but hanging on can only make things worse for them. □

Problems with anti-clotting drug

- First-year sales only half of projected values
- Accusations of withheld information
- Advantages over competitors not clear-cut

Berkeley

STORM clouds have gathered over Genentech's new clot-dissolving drug, Activase (the company's trade name for tissue plasminogen activator, TPA). News that TPA sales are lower than expected have the South San Francisco company's stock price plunging. Now, questions have arisen over the ethics of the company's promotion and testing of the drug.

Seven class-action lawsuits have been filed against Genentech, on behalf of stockholders claiming that the company withheld adverse information about TPA's promise while company insiders sold off \$24 million in shares. In addition, staff of a congressional committee have discovered that at least fifteen of the researchers participating in trials of TPA owned Genentech stock.

Activase was greeted with fanfare when it was released last November, and sales of the drug amounted to \$58 million by the end of December. Although the alternative drug, streptokinase, had been available for some time, TPA's expected advantages, such as greater clot specificity and fewer allergic or bleeding side-effects, promised to help it to obtain a large share of the thrombolytic drug market, despite its price of \$2,200 a dose, eleven times that of the less glamorous streptokinase.

With first-year sales projected at about \$180 million, TPA will be the most successful first-year drug ever. But sales have nevertheless fallen short of predictions roughly twice as great, and the price of Genentech stock has fallen by a half since the beginning of the year. Company officials have blamed TPA's slower-than-expected start on physicians' confusion over trial data rather than company over-enthusiasm for its product.

Indeed, there is little agreement among cardiologists about the meaning of the various trials of TPA and its competitors, indicating that TPA's advantages are less clear-cut than enthusiasts expected. The lack of a large head-to-head mortality trial comparing TPA and streptokinase complicates the comparison.

The result is that cardiologists find support for opposing views by dipping into the same pool of data. TPA supporters say the drug is significantly better at clearing arteries when administered during the critical first four hours after a heart attack, and carries a lower risk of recurrent blockage or internal bleeding. But TPA

opponents use the same studies to conclude that the drug is not significantly better than streptokinase, and carries higher re-blockage and bleeding risks.

The recent lawsuits accuse Genentech of withholding the drug-trial ambiguities from investors, but Denise Gilbert, a market analyst with Montgomery Securities in San Francisco, says the charges are unlikely to stick, because analysts knew of the trial results as soon as the medical community.

But in spite of the confusing results, many analysts believed Genentech's insistence that TPA would be the drug of choice for heart attacks. Although reluctant to fault Genentech for what they



describe as a "brilliant" marketing campaign, some now feel manipulated or misled.

Genentech reported last April that TPA use was growing, and had reached 2,000 doses a week, up from 1,000 at the beginning of the year. Then, according to Gilbert, they stopped talking about patient numbers. In July there was news that sales were lower than predicted, but that was attributed by company officials to over-stocking by hospitals and suppliers when the drug first came out, and analysts such as Gilbert expected sales to bounce back as the surplus shrank and usage grew. So they were shocked to learn in September that TPA usage had remained constant since April. Two weeks later, Genentech announced suspension of TPA production, because of a surplus created during the first half of the year.

Gilbert questions why the company's July announcement did not mention that TPA usage had been flat since April. Moreover, it cannot have been constant at the claimed rate of 2,000 patients a week,

she says, because the company is now saying the total number of patients treated with TPA since last November is 50,000, a number more consistent with a usage rate of 1,000–1,200 a week.

Genentech denies intentionally making over-optimistic sales forecasts, and has refused to comment on the lawsuits, beyond a statement that they contain "numerous factual inaccuracies" and that the company plans to defend them "vigorously".

Some observers of the biotechnology industry say the over-optimism was generated as much by enthusiastic cardiologists as by Genentech's marketing. But evidence gathered by the staff of the House of Representatives government operations subcommittee on human resources and intergovernmental relations revealed that at least 15 researchers participating in TPA trials owned stock in Genentech. The subcommittee staff investigation has also revealed that Genentech offered options to buy sizeable amounts of its stock to three additional researchers, including one who was a member of the executive committee of the Thrombolysis in Myocardial Infarction (TIMI) trial sponsored by the National Institutes of Health.

At a human resources subcommittee hearing on 29 September, a member of the TIMI monitoring committee testified that committee members seemed strongly biased towards TPA before the trial even began. Victor Marder, co-chief of haematology at the University of Rochester, said an atmosphere of urgency drove the committee to begin a second phase mortality trial before results from a preliminary trial comparing TPA to streptokinase had been adequately analysed. He said committee members seemed so convinced that TPA was better that they dropped streptokinase from the more extensive mortality phase of the trial, thus eliminating an opportunity for a direct comparison of the drugs.

TPA supporters staunchly argue that it is more effective and safer than its competitors. But some, weary of the debate, say either drug can save lives, and placing politics ahead of medicine has scared away conservative cardiologists, so that only a fraction of the several hundred thousand heart-attack victims who stand to benefit from the drugs are receiving them. "The death rate from acute MI has been cut in half by thrombolysis when implemented early", said Burton Sobel, director of the cardiovascular division at Washington University, and a consultant to Genentech. "We have a truly phenomenal advance, and it's not being used widely in 1988."

Marcia Barinaga

1988 Nobel prizes announced for physics and for chemistry

Washington

THIS year's award of the Nobel prize for physics recognizes fundamental experiments which belong almost to the pre-history of the fast-moving field of high-energy physics. The recipients of the prize are Leon Lederman, Melvin Schwartz and Jack Steinberger, all Americans, and the work cited is their invention, in the early 1960s, of a way to generate a neutrino beam in the laboratory and their subsequent discovery of the muon neutrino.

Any surprise at the award derives from its belatedness rather than the stature of the winners. Lederman, Schwartz and Steinberger, according to the gossip that passes around high-energy physics departments, had appeared on the short-list numerous times, but with prizes having already been awarded for more recent discoveries, there was a general feeling that the three had run out of chances. Their win this year has been applauded by others in experimental high-energy physics as deserved recognition of basic work.

In the late 1950s, theoretical understanding of the weak interaction (responsible for radioactive beta-decay) was limited by the inability to do experiments in which its effects could be separated from those of the strong and electromagnetic interactions. At Columbia University, the three physicists jointly devised a way to produce and detect a beam of neutrinos, particles influenced by the weak interaction and no other.

By smashing an energetic proton beam into a beryllium target, they first created a stream of nuclear fragments containing a large proportion of pions. Each pion then decays in flight into a neutrino and another charged particle. The second, equally important step was to separate the neutrinos from the nuclear debris. This they achieved by erecting a massive steel wall, made of scrap steel from battleships, behind which they built a 10-ton spark chamber to detect secondary charged particles produced by collisions between neutrinos and atoms in the chamber.

The unexpected bonus of this experiment, whose unprecedented size made it the precursor of the vast experimental setups now routinely built at accelerators, was the discovery of two different neutrino types. At the time only one neutrino, produced in accompaniment with an electron in beta-decays, was recognized, but in the new experiment it was found that the charged particle that accompanied the spark-chamber detection was more often a muon than an electron, and a series of experiments soon established that the muon neutrino was a species distinct from the previously known electron neutrino.

This was a first indication of the now-standard recognition that elementary particles, including quarks, neutrinos and their accompanying charged partners, fall into three parallel but separate families.

Schwartz now runs his own computer communications company in Mountain View, California, but Lederman, director of the Fermi National Accelerator Laboratory in Illinois, and Steinberger, a physicist at CERN, Geneva, are still active in high-energy physics. Lederman is an active supporter of the Superconducting Super Collider, the proposed 50-mile-circumference proton-proton collider. Under his tenure, Fermilab has run a successful programme that brings high-school students into the laboratory for summer courses.

David Lindley

London

ELUCIDATION of the structure of a protein complex with its "body immersed in a lipid bilayer and its head and legs in water" has won Hartmut Michel, Johann Deisenhofer and Robert Huber this year's Nobel prize for chemistry. The structure is the membrane-bound photosynthetic reaction centre of purple bacteria, which converts the energy of a photon into the potential to synthesize organic molecules.

Not only is the solution of the structure an impressive technical feat, but reaction centres are one of the most important protein structures in biology. The complex of different proteins within the centre is the

heart of the photosynthetic reaction.

Crystallization of membrane proteins was a goal that had been sought by many, but Hartmut Michel, working in the division of Dieter Oesterhelt at the Max Planck Institute for Biochemistry at Martinsried, Munich, made a crucial breakthrough in 1981 by devising a method to produce large, well-ordered crystals of the reaction centre of the purple bacterium *Rhodospseudomonas viridis*.

The year before, two other membrane proteins had been crystallized for the first time, but technical problems prevented the arrangement of the atoms within the crystals from being determined. It had previously been difficult to isolate membrane proteins because the detergent necessary to dissolve the protein away from the lipid cell membrane tended to denature the protein itself. But a new detergent, octylglucoside, allowed three-dimensional crystals of two protein complexes to be made: porin, a protein that forms pores in cell membranes; and bacteriorhodopsin.

The race was on to make three-dimensional crystals of a membrane protein large enough to resolve its structure in atomic detail using X-ray crystallography. Michel was the first to make such crystals, not of any of these proteins, but of the bacterial reaction centre.

Once Michel had obtained large three-dimensional crystals of the reaction centre, he began his fruitful collaboration with Robert Huber, head of the laboratory, and Johann Deisenhofer, a staff scientist.

Their first results, published in 1984

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Top, Lederman (left), Schwartz and Steinberger, discoverers of the muon neutrino. Bottom, Deisenhofer (left), Huber and Michel, first to solve the structure of a membrane protein.

(see below) showed that the central portion of the reaction centre consists of a 50-Å-long cylinder and that the whole complex has a diameter of 140 Å.

In 1985, Michel and his collaborators reported the molecular structure of the whole reaction centre obtained from analysing the X-ray diffraction pattern of the crystals, emerged with great clarity. The structure confirmed predictions about how energy transfer works as well as revealing unusual features.

In the photosynthetic reaction, light energy is absorbed by electrons in pigment molecules surrounding the reaction centre complex in the membrane. The energy of the photon is rapidly transferred to the reaction centre, exciting the bacteriochlorophyll which then gives an electron to an acceptor. Earlier this year, G. Fleming *et al.* (*Nature* 333, 190; 1988) identified the acceptor as the bacteriopheophytin in the reaction centre complex. This electron transfer produces energy which is used by the complex to oxidize water, releasing oxygen. The other products are hydrogen and electrons, which reduce carbon dioxide to organic compounds.

Since 1985, techniques such as site-directed mutagenesis have revealed more details about how the reaction centre works. And although the Nobel-prize-winning work holds out great promise for the elucidation of the structures of other membrane proteins, these hopes have not yet been realized. None of the membrane protein complexes crystallized in 1980 has yet been seen at atomic resolution.

The Nobel-prizewinning team separated earlier this year; Michel moved to the Max-Planck Institute for Biophysics at Frankfurt, and Deisenhofer to the Howard Hughes Medical Institute at Dallas.

Maxine Clarke

■ An editor's nightmare is to reject a Nobel-prizewinning paper. Rejected authors not infrequently promise that the nightmare will come true. Hartmut Michel did not issue such a warning when we declined to publish his report of the successful crystallization of the photosynthetic reaction centre, saying that we looked forward to the time when the crystal yielded structural information.

When the first such information — on the bacteriochlorophylls and other prosthetic groups in the complex — became available, Michel and his colleagues published it in the *Journal of Molecular Biology* (180, 385; 1984). But when the structure of the protein subunits emerged, allowing the complete description of the reaction centre, *Nature* was delighted to be able to offer to publish it (318, 618; 1985).

Rejection of Hans Krebs' discovery of the tricarboxylic acid (or Krebs') cycle, a pivot of biochemical metabolism, remains *Nature's* most egregious error (as far as we know). □

Soviet academy rights one wrong but stands accused by Legasov

London

ACADEMICIAN Andrei Sakharov was last week elected to the presidium of the Soviet Academy of Sciences, by a vote of 234 to 84. His election took place during a three-day general meeting of the academy, devoted to the role of science in *perestroika* and the election of a younger leadership for the academy. Before the meeting, academy president Dr Gurii Marchuk said that a 50 per cent 'renewal' of the leadership was planned, to bring in young scientists, including five new vice-presidents.

In terms of the Gorbachev era, however, Academician Sakharov, at 67, hardly qualifies as 'young' and his well-deserved, even if belated, election to the presidium seems rather a redressing of an old injustice. Yu. A. Osipyan (physics), O.M. Nefedov (chemistry), R.V. Petrov (biology), N.P. Laverov (Earth sciences) and V.N. Kudryavtsev (social sciences) were eventually elected vice-presidents.

Marchuk's keynote speech stressed the

Poor return from India's science effort

New Delhi

New statistics from the Indian Department of Science and Technology explode the myth that India comes third in national rankings of the number of working scientists. Although India boasts an estimated four million people with training in science, the report shows that a mere 85,309 are working on research and development. Of this number, just 4,375 are women.

Among the 241,000 persons employed at Indian research institutions, only one in four is a researcher. The remainder are administrative and auxiliary staff.

The report estimates that India spent 28,655 million rupees (£2,000 million), or 1.1 per cent of its gross national product, on research and development in 1986-87, a 29 per cent rise over the previous year. About a third of the budget was spent on research activities related to defence. Industrial production and energy sectors received 12 and 10 per cent, respectively and transport and communications around 4 per cent.

The report notes that industrial research and development has increased, largely due to liberal financial incentives from the government. The private sector provides close to 20 per cent of expenditure. But the report warns that output is not commensurate with investment. The number of patents filed by Indian citizens has been declining since 1977. K. S. Jayaraman

role of science in developing the concept of a state that would reflect "all the values of the socialist order and, first and foremost, concern for the human being".

Yet, ironically, only a few days before the meeting, an article in *Pravda* suggested that conservatism within the academy may have played a significant role in the suicide of Academician Valerii Legasov last April. According to the inquest, Legasov took his life in a state of depression, while popular rumour associates his death with the high doses of radiation he received while overseeing the Chernobyl clean-up. *Pravda*, however, suggested that the academy old guard added to his worries. In spring 1987, when the academy voted on the appointment of a new director of the Kurchatov Institute of Atomic Energy, Legasov, although chief deputy director, was voted down by 129 to 100, a disappointment which, *Pravda* said, "shattered" him.

His efforts to publicize the neglect of safety training and general corner-cutting that lay behind the Chernobyl catastrophe seemed to have no effect. Fellow scientists, *Pravda* said, dismissed his views with phrases such as "Legasov does not follow the principles of Kurchatov". He was told that he would receive the title of Hero of Socialist Labour for his work at Chernobyl, but then learned that no one from the Kurchatov Institute would be so honoured, as the scientists there were considered to be to some extent responsible for what happened.

Finally, on 26 April 1988, the second anniversary of Chernobyl, a plan worked out by Legasov for reactivating Soviet chemistry after the 'stagnation' of the Brezhnev era was rejected by senior academicians with the command that "We won't let a mere lad lead us". (Legasov was 52.) The following day, he committed suicide. His article on Chernobyl "It is my duty to speak out..." finally appeared in *Pravda* a month after his death.

Legasov's death may be construed as over-reaction to professional disappointment while in an abnormal state of health. Nevertheless, the *Pravda* account gives an unusual insight into the problems of *perestroika* within the academy. The central committee of the Communist Party of the Soviet Union has passed a resolution to commemorate Legasov's name (by calling streets, schools and possibly a research vessel after him) — and the recent Seventh World Conference on Hydrogen Energy in Moscow opened with a minute's silence in his memory. But the best tribute to his memory, *Pravda* concluded, would be the implementation of his ideas, on energy, ecology, safety and *perestroika*.

Vera Rich

US anger over accusations of trafficking in infant organs

Washington

THE US State Department has reacted with outrage to a resolution passed by the European Parliament last month that accuses the United States of trafficking in organs from human infants. US officials maintain that the 'baby parts' issue is part of a Soviet disinformation campaign, and has no basis in fact.

Todd Leventhal, a policy officer at the US Information Agency, says the first occurrence of the organ trafficking story appeared in a Honduran newspaper on 2 January 1987, attributed to Leonardo Villeda Bermudez, the former secretary general of the Honduran Committee for Social Welfare. The story at the time was that there were special fattening houses where babies were raised, and then their organs harvested for sale in the United States. Leventhal says Bermudez denied the reports attributed to him the next day, but his denial is never mentioned in subsequent stories. Similar stories have surfaced involving Guatemala.

The parliament resolution, introduced by a French Communist member, "condemns in the strongest terms practices of this kind", and calls on governments to hold inquiries into the affair.

Richard Schifter, Assistant Secretary of State for Human Rights and Humanitarian Affairs, in a letter dated 8 October to Karel de Gucht, chairman of the parliament's human rights subcommittee, expresses outrage over the "gullible acceptance by the Parliament of a totally untrue statement". US officials admit that there do exist illegal adoption services in Latin American countries that try to sell babies for adoption in the United States. Leventhal says the Soviet Union has used this fact to insinuate that the baby's organs are being marketed as well. He says the baby parts campaign is reminiscent of a similar campaign that claimed that AIDS was the result of a US biological weapon experiment gone awry.

US irritation over the AIDS campaign prompted the cancellation of several visits between US and Soviet health officials to discuss AIDS issues. Ironically, just days after Schifter sent his letter, the US Institute of Medicine and the Soviet Academy of Medical Sciences signed a five-year protocol for joint research programmes. The new protocol calls for cooperation on research on AIDS, environmental release of radiation, alcohol and drug dependency, and polio.

Joseph Palca

Conflict of interest over Harvard drug

Boston

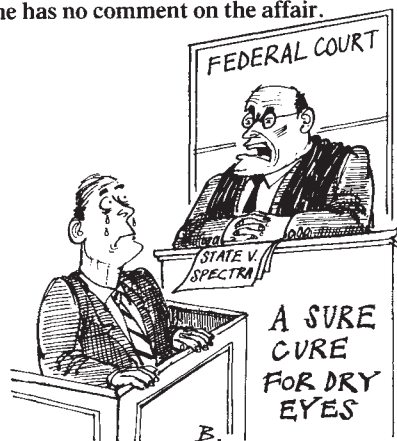
IN an unusually candid statement released last week by Harvard Medical School, Dean Daniel C. Tosteson acknowledged that a "significant conflict of interest occurred" in a drug study undertaken by a former Harvard researcher. The study, involving an experimental ophthalmic ointment, occurred three years ago, but it has been the focus of continuing investigation since then and has heightened concern about the effect of financial interests on biomedical researchers.

The researcher involved, Scheffer C. G. Tseng, and several of his colleagues and relatives, have apparently made large profits from the therapeutic promise of the ointment, using for the purpose an arrangement with a small company they set up called Spectra Pharmaceutical Services, Inc.

The drug, called Tretinoin, is an ointment containing vitamin-A. The medication was purported to cure dry-eye conditions caused by a family of ophthalmic diseases, including *keratoconjunctivitis sicca*. Ten million people are estimated to be afflicted with dry eyes in the United States alone.

Harvard says Tseng violated university research guidelines by administering the

drug in clinical trials without proper authorization. Tseng, who left Harvard more than two years ago for the University of Miami Medical School, refused several requests for an interview, and says he has no comment on the affair.



Harvard's statement about the incident says that Tseng failed to make "timely and complete disclosures" about changes in his study and his financial interest in the drug. The university also concedes that Tseng received "insufficient" supervision and that his study was run "without sufficient safeguards to protect [against] potential bias".

Seth Shulman

NIH delay first gene transfer

Washington

JAMES Wyngaarden, director of the National Institutes of Health (NIH), has decided to defer approving the first human gene transfer experiment. Although the NIH Recombinant DNA Advisory Committee (RAC) approved the experiment at a meeting earlier this month (see *Nature* 335, 577; 1988), Wyngaarden in a letter dated 18 October to RAC chairman Gerard McGerrity wrote that he wanted the protocol reevaluated on the basis of "unresolved questions" raised by the RAC's human gene therapy subcommittee and the NIH Institutional Biosafety Committee.

The experiment, proposed by W. French Anderson of the National Heart Lung and Blood Institute and R. Michael Blaese and Steven A. Rosenberg of the National Cancer Institute, involves putting a bacterial antibiotic-resistance gene into human lymphocytes specially selected for their ability to attack cancer tumours.

These 'tumour infiltrating lymphocytes' are grown in culture, activated with interleukin-2, and then injected into the patients from whom they were originally obtained. The bacterial gene provides no direct benefit for the patient, but will enable researchers to track how well the lymphocytes re-infiltrate their target tumour.

The gene therapy subcommittee was established to evaluate the details of the Anderson, Blaese and Rosenberg protocol, and at a meeting on 29 July it asked for additional data about the experiment. On 29 September, the subcommittee met and concluded that it had not yet received sufficient data to reach a decision on the proposal. But five days later, Anderson presented the additional data to the full RAC, and the experiment was approved. Anderson explained at the time that he had not provided the subcommittee with complete, written answers to its questions because he felt that to do so might jeopardize his prospects for subsequent publication in a scientific journal.

In his letter to the RAC, Wyngaarden says that "although this does not constitute gene therapy *per se*, gene transfer must be regarded as a prelude to such experiments and thus subject to the identical review process". Wyngaarden asks that the subcommittee re-evaluate the protocol quickly. No timetable has yet been established for this review.

In addition to the re-review by RAC and its subcommittee, the protocol must also be approved by the institutional review boards of the two NIH institutes involved, and the NIH Institutional Biosafety Committee.

Joseph Palca

Herstmonceux castle sold but not to the highest bidder

London

HERSTMONCEUX castle, for 40 years the home of the Royal Greenwich Observatory, was sold last week to a British property developer for between £7 and £10 million, more than double original estimates of the site's value. The exact price paid has not been disclosed.

Possible plans for the site, which includes parkland, an observatory and offices, as well as the castle, now include a conference centre, hotel, private residences, a golf course and a science centre for the public. James Developments, the new owner, outbid a group of researchers who planned to transform the site into a science centre. The group, led by Professor Richard Gregory of the University of Bristol, raised £7 million with the help of Irving Abrams, a Canadian businessman.

When that was beaten, they made an increased offer of £14 million. But the rules of the sale meant that that bid, made after the original closing date for offers, could not be accepted. Gregory is now negotiating with the developer with a view to securing part of the site for his plans.

Ian Tegg, managing director of the developer, says he is looking for "bright people with bright ideas" to determine the future of Herstmonceux, and says there is interest in the site from the United States, Canada and Japan as well as Britain. Whether the observatory will continue to be used or will become a museum has not yet been decided. Staff move to a new building in Cambridge in 1990. The move will cost about £7 million and will be financed by the sale. Funds remaining will go to the Treasury. **Christine McGourty**



Herstmonceux castle — into an unknown future.

French government already under attack

Paris

THE socialist French government is experiencing its most troubled period since it was elected last spring. By refusing to give in to demands for pay rises in the public sector, the government has angered teachers and nurses whose interests the socialist party claims to represent.

While thousands of public sector employees, especially nurses, teachers and those in nationalized industries, protested in the streets of Paris last week, the government faced the possibility that its 1989 budget would be thrown out if 'centrists' joined with the right-wing opposition. But although Prime Minister Michel Rocard is accused of squandering the nation's new-found wealth, the budget increase in research spending has been applauded on both sides of the assembly.

The government has been accused of 'anti-Europeanism' by the right-wing

opposition because of its resistance to tampering with the tax structure. To fall in line with other European countries from 1992, France will have to reduce its rates of value-added tax (VAT). But increased revenue from VAT during 1987 was an important factor in the new government's ability to reverse its predecessors' demands for belt-tightening in higher education and research. Now it looks as though the finance minister, Pierre Bérégovoy, will have to reduce VAT on luxury goods, but Rocard hopes to stall the European project to introduce a uniform fiscal structure.

The most delicate series of debates will be held this week as the budgets of individual ministries are scrutinized. It is likely that the government will leave the most controversial ministries to last, including education, in the hope of winning centre support in time for a single vote to approve the budgets *en bloc*. **Peter Coles**

Telescopes get the green light

Washington

THE University of Arizona will have a new observatory in its backyard, thanks to a law passed in the final days of this congressional session. As part of the Arizona-Idaho Conservation Act of 1988, the university will be allowed to build three telescopes on one of the 11,000-foot-high peaks of Mount Graham, in southeastern Arizona.

The law ends an often contentious four-and-a-half-year evaluation and environmental assessment of the university's plans by the US Forest Service, the federal agency charged with managing the mountain. The new legislation will make it extremely difficult for opponents of the observatories to delay or prevent their construction.

The law allows three telescopes — the Columbus Project, a Vatican telescope and the Max Planck sub-millimetre telescope and their accompanying support facilities — to be built on 24 acres of Emerald Peak, subject to conditions outlined by the US Fish and Wildlife Service (see *Nature* 334, 645; 1988).

The fate of four additional telescopes sought by the university will depend on how the endangered Mount Graham red squirrel fares during the initial construction phase, according to Patty Lynch, legislative assistant for Senator Dennis DeConcini (Democrat, Arizona).

The new legislation on the observatories was a late addition to a bill that authorized several land transfers, including the acquisition by the federal government of 100,000 acres of Florida Everglades wetlands.

Scientists for the Preservation of Mount Graham, a group of about 100 biologists, many of whom had worked on the mountain, opposes any observatory development because the peaks represent a unique "desert sky island" of sub-alpine spruce and fir.

As part of the law, the university must develop and fund a conservation management plan in conjunction with federal agencies. "The University of Arizona will be a responsible steward", said Allan Beigel, vice-president for university relations. "We see ourselves as having linked hands with responsible environmentalists."

But opponents of the university's plans vow to continue the fight. "We're going to try to stop the addition of four more telescopes", says Paul Hirt, local Sierra Club representative and spokesman for the Coalition for the Preservation of Mount Graham. He also vowed that the coalition would be a watchdog to make sure the university sticks to its authorized plans. **Elizabeth Pennisi**

Generic drug manufacturers lose patent protection battle

London

In the conflict between the originators of new drugs and the manufacturers of generic copies of the originals, the originators are about to score a significant victory. Next week a change in patent law will give certain drugs, including some of the world's most successful, an extra four years of protection in Britain from generic competitors. The originators say this will strengthen the industry as a whole; they argue that better patent protection increases confidence in the industry and as a result increases investment. And they claim that increased patent life is necessary to ensure recovery of the increasing costs of research. Although it may damage the generics industry in the short term, in the long it will be beneficial, they say.

But independent companies claim that

Use of fetal tissue deemed acceptable

Washington

A SPECIAL panel convened to answer questions about research using human fetal tissue last week concluded that although "it is of moral relevance" if the tissue is obtained from induced abortions, nevertheless "the use of such tissue is acceptable public policy".

The Human Fetal Tissue Transplantation Research Advisory Panel was formed at the request of the US Health and Human Services Assistant Secretary Robert Windom, who placed a moratorium on federal funding for fetal tissue research using material from induced abortions pending the outcome of the panel's deliberations. The panel began work last month, but recessed after two days of discussion without reaching agreement on precise language for answering the secretary's questions.

The panel took pains not to take a moral stand on the contentious issue of abortion, instead concluding that so long as abortion was legal, then cadaverous tissue obtained from aborted fetuses could legitimately be used for research. Opponents of abortion had argued that expanding the use of fetal tissue would encourage women to have abortions, but the panel concluded that there was no evidence that use of fetal remains for research had had "a material effect on the reasons for seeking an abortion in the past".

The panel's recommendations will be presented to James Wyngaarden, director of the National Institutes of Health, at the directors' advisory meeting on 1 December, and Wyngaarden will transmit them to the assistant secretary. Joseph Palca

the industry will be severely hit; some smaller companies may not survive. In addition, it is claimed that between £50 and £100 million may be added to the National Health Services medicines bill. And despite an independent review which said the repeal of the proviso was not justified on the grounds of increased investor confidence and which documented the substantial extra cost to the National Health Service, the government decided for the repeal. It will remove a proviso in the 1977 patents act which has irritated the mainstream pharmaceutical companies for several years. When patent protection for drugs in Britain was increased from 16 to 20 years in 1977, nobody noticed a proviso which, for certain drugs, effectively took back the extra four years the act granted.

It meant that for products under patent protection at the time of the act, and with more than five years of patent life to run, other companies could apply for a licence to use the patent during the extra four years granted, providing that royalties were paid to the originator. Patents filed after introduction of the new act were unaffected by that proviso, so it was destined to become obsolete in the 1990s, leaving a four-year 'product vacuum' as it was phased out. The imminent repeal of the proviso, which was hotly debated earlier this year, brings this 'vacuum' forward by five years. The generic equivalents of Zantac (Glaxo) and Zovirax (Wellcome), for example, were due to become available in 1993 and 1991, respectively. Now availability will be delayed until 1997 and 1995. A study by Martin Paltnoi Associates, a British consultancy which specializes in drug patents, says that about 50 drugs will be affected.

The Association of the British Pharmaceutical Industry has been lobbying for the repeal for several years. It claims that poor patent protection in Britain makes it relatively unattractive compared to its international competitors. In Britain, about 12 years of patent life is taken up in getting the product to the marketplace, leaving only eight years of protection, of which half were lost for many products as a result of the licence proviso.

Canada and the United States are increasingly seen to be more attractive to investors because of their patent law, say the British manufacturers. In Canada, a law passed last year allows drug patent holders 10 years of exclusive marketing rights. In exchange, the industry promises to double research spending within 10 years. In the United States, a law passed in 1984 allows manufacturers to apply for an extension of the patent life of 17 years to

AIDS reagent repository opens

Washington

AFTER years of preparations, an AIDS reagent repository containing deposits from AIDS researchers world-wide has been opened by the US National Institute of Allergy and Infectious Diseases (NIAID). The repository contains over 100 reagents, virus stocks, cell lines and clones drawn from contributions solicited from nearly 1,500 researchers.

The repository was established in April of this year through the efforts of John McGowan, chief of the Developmental Therapeutics Branch of NIAID's AIDS programme. McGowan says the two aims of the repository are to make reagents freely available to draw in new researchers to study AIDS, and to offer standardized protocols and reagents so that results can be compared between laboratories.

The NIAID repository is part of the World Health Organization AIDS reagent repository programme, and is collaborating with similar repositories in Paris and London. The repository's reagents are free to industry, government and academic researchers, to be used for research purposes only. Carol Ezzell

Magellan damage

Washington

THE Magellan spacecraft, scheduled to be launched towards Venus in April next year, suffered only minor damage when an electrical fire occurred on 17 October as it was being set up for testing at the Kennedy Space Center, Florida. The fire was caused when wires connecting an external battery touched each other, and was extinguished within a minute. Project scientist Steve Saunders said last Monday that after visible damage had been cleaned up, tests showed that the spacecraft's signal processing units were unscathed. The only concern, he said, was the loss of ten days' time from the launch schedule, although this does not immediately jeopardize the launch date. David Lindley

compensate for time taken seeking regulatory approval. This law is presently being tested; it is being disputed whether time spent on research and development can count as time spent seeking regulatory approval. On the introduction of that law, the generic sector was compensated by being permitted to carry out research during the patent life of the product.

Jorge Goldstein, a US patent lawyer, described an unexpected spin-off from that law: aeroplane manufacturers are now claiming that patent lifetime of their products is absorbed by the time spent seeking approval of the Federal Aviation Authority. But such claims are unlikely to succeed, he says. Christine McGourty

Problems of small countries in linking research to industry

Ottawa

SMALL countries may have special problems in finding a way to link basic research to the industrial research and development needed for economic prosperity, according to a new report*.

The thrust of basic research in small countries will "be primarily determined by the links that scientists form with their peers in the international basic research community". In contrast, in those with a strong industrial science base, some basic research will be naturally linked to the industrial innovation process.

At one pole are countries such as Canada and Australia whose industries carry out very little research and exert almost no influence on basic research. They, and countries such as Britain which have the same historical tradition, share the problem of being unable to link science (based in universities) effectively to technology (based in industry). But the problems are worse for the smaller countries where it is very difficult to start technological growth in industry.

At the other pole stands Japan which has "come to science from engineering and technology". Basic research laboratories are being built up in industry, rather than in the universities, and so can directly enhance effective applied research.

The report argues that innovation cannot be produced by putting pressure

on the universities "to act as surrogates for what industry should be doing". Small countries have to come up with new institutional arrangements to "pool resources within and outside their country" and learn how to maintain the integrity of the pyramid of basic, applied and developmental research within industry.

They face stiff competition. The combined research and development budget of Japan's six big electronics companies is almost as big as the entire research and development budget of Italy and bigger than those of any Western nation excluding the United States, Germany, France and Britain.

Small countries can, however, do well, as Sweden and Switzerland show. Canada is given a fair chance of improving its miserable position at the bottom of almost every measure of research and development performance in the OECD league table. The economy is expected to do well for the next five years and will provide a stable period for change. The new principles needed include trust and consensus among all the different political, industrial and research actors and an agreement that major priorities for science-based innovation must be industry-driven.

Alun Anderson

* *Innovation and Canada's Prosperity: The transforming power of science, engineering and technology* (Canadian Institute for Advanced Research, 179 John Street, Suite 701, Ontario M5T 1X4, Canada).

Aflatoxin contamination of US corn

Washington

THE severe drought in US corn-producing states last summer not only reduced this year's corn crop but also gave a foothold to crop contamination by a mould which produces the most carcinogenic natural product known: aflatoxin. An Iowa State University survey released last week of aflatoxin contamination of corn in Iowa grain elevators revealed that over one-third has levels of aflatoxin toxic to humans. Illinois and Indiana have reported similar levels, and aflatoxin has also been detected in corn grown in Wisconsin and Minnesota.

Aflatoxin is produced by roughly half of the natural population of *Aspergillus flavus*, a ubiquitous mould which invades only corn that has been stressed by high temperatures and low moisture. Aflatoxin contamination is relatively common in southern states such as Georgia and Alabama, but troubling levels have not been found in midwestern corn crops since 1983.

The US Food and Drug Administration (FDA) outlawed the interstate shipping and foreign export of corn containing aflatoxin levels greater than 20 parts per billion (10⁹), the highest amount reckoned safe for

human consumption. But beef cattle are able to metabolize larger amounts of aflatoxin without tainting the meat, so the FDA raised the maximum levels allowed in feed earlier this month. Under the new rules, mature cattle can eat corn containing up to 300 parts per billion aflatoxin, swine can consume feed with 200, and breeding livestock are allowed 100. Because dairy cattle can pass the toxin along in their milk, corn fed to dairy cattle must contain no more than 20 parts per billion aflatoxin.

But critics of the new FDA regulations say that dividing up the corn crop by aflatoxin level is not a workable solution. Local grain elevators routinely screen corn delivered by farmers by passing it under an ultraviolet light to detect mould, but the elevators are not capable of assaying corn specifically for aflatoxin. Most grain is tested for aflatoxin just before it is bought and shipped, and large buyers of corn may be reluctant to purchase shipments which have restricted uses. Charles Hurburgh, who headed the Iowa survey, says "the scrutiny is tightening by the day" and that market pressures will keep higher-aflatoxin corn from being sold. Carol Ezzell

Delhi science centre

New Delhi

A SCIENCE and technology centre for non-aligned countries is to be opened in New Delhi early in 1989. The centre has been under discussion for two years.

Thirty-one non-aligned developing nations, including China and Pakistan, have agreed to participate. Others are expected to join before the governing council meeting in February 1989 that will settle remaining issues of funding and organization.

A spokesman of the Indian Department of Science and Technology said one aim of the centre is to pool the talent and resources available within the non-aligned nations to promote collective self-reliance and reduce dependence on the West.

A recent meeting of scientists from 25 countries identified five areas for joint ventures in research and development: information and telecommunications, microelectronics, computer software, biotechnology and renewable sources of energy. The centre proposes to set up common databanks and information networks covering these fields. There are also proposals for computer time-sharing and exchange of scientists and students.

The centre will also try to develop collective strategies to deal with such issues as the dumping of toxic wastes by industrialized nations, and the impact of advanced technologies on the economies of developing countries.

K. S. Jayaraman

South Africa avoids expulsion again

Oxford

SOUTH Africa successfully staved off expulsion from the International Atomic Energy Agency (IAEA) yet again at the organization's annual general assembly in Vienna last month. A group of African countries, led by Nigeria, were forced to back down from a bid to oust South Africa from the agency, after the Soviet Union and other Eastern bloc states failed to support the move for the second year running.

The group said that it had been forced to withdraw its resolution until next year's conference, after Soviet delegates had indicated that they believed South Africa would soon sign the Nuclear Non-Proliferation Treaty (NPT). In August, a South African delegation was urged by British, US and Soviet representatives to do so, and in September a second delegation met the representatives of 30 countries to get clarification on the terms of the treaty.

A spokesman for the South African Department of Foreign Affairs said that an in-depth study of the implications of signing the treaty had been accorded priority status. Until this was completed, he claimed, the country would not commit itself to signing the NPT. Michael Cherry

Scientists' attitude to weapons

SIR—*Nature* is an influential journal among professional scientists, and for this reason I would like to see some dialogue about the morality of weapons science in your Correspondence columns. This is an issue of professional importance to some of your readers (those in the weapons industry) and of life or death importance to all.

We must ask ourselves how to stop the military madness which infects economies, societies and science. For centuries, scientists and engineers have been employed by governments to move troops across rough terrain or maintain the health of troops in hostile environments. These innocuous military advances have produced many benefits for our society such as air transport and tropical medicine. But within the lifetime of many of your readers, science has evolved as the agent of death for the innocent. Unlike our ancestors' inventions, nuclear, chemical and biological weapons are designed for genocide. The superpowers have manipulated scientists to do their bidding by appealing to patriotism, paranoia and purse-strings. I believe these scientists and engineers must be accountable for their actions.

No president, secretary, dictator or monarch knows how to build a nuclear weapon or a recombinant pathogen. They only know how to manipulate the people who have the know-how. Hitler did not design a gas chamber, a patriotic engineer did the dirty work. V2 rockets sprang from the minds of scientists doing their duty for their country. Even some members of the medical profession, in spite of the hippocratic oath, tortured and murdered civilians in the paranoia of Nazi Germany. History can teach us only what we will allow ourselves to learn and learning begins with tough questions. Who made the chemical weapons that killed civilians along the Iran-Iraq border recently? Are they pleased with the results? Do these chemists condemn their historical counterparts or emulate them under a different guise? Who are the engineers who develop armaments which so effectively kill 'rebels' all over the world (on both 'sides')? Do they defend their actions by claiming "I only build them I don't use them!"? Will I be a victim of my fellow scientists' creation? Do the scientists and engineers who develop these weapons expect their respective governments to shoulder the moral responsibility?

As scientists and human beings, we have an obligation to question our motives and make moral decisions about our involvement in the military-industrial complex. The minds and money wasted on military applications are staggering and could be put to constructive use. It is too late to keep the secrets of mass destruction from

falling into the wrong hands, but it is not too late to teach our graduate students and peers that it is wrong to research and develop more efficient means to murder and destroy. Perhaps it is time that the scientific community organized to oppose those forces that cause us to be technocomplices to murder.

JAMIE LOVE

*Department of Biochemistry and
Molecular Biology,
Louisiana State University Medical
Center,
1901 Perdido Street, New Orleans,
Louisiana 70112-1393, USA*

Bad theology

SIR—It was good to be reminded by Euan Nisbet that the problem with so-called 'creationism' is not just that it is bad science, but also that it is bad theology (*Nature* 334, 575; 1988). Nisbet's review also underlines the fact, sometimes lost in the rather murky and polarized 'creation/evolution' debate, that the great majority of scientists who are also Christians do their science very happily within the paradigm of evolution. The fact that we are a rather 'silent' majority, should not detract from the historical evidence showing not only that evolution was accepted remarkably quickly by most church leaders of the last century¹, but also that many well-known Christians at the time were active proponents of Darwin's theory².

The creationists' insistence that the Bible should be read as a kind of scientific textbook is an idea that was heavily criticized by the leading figures of the seventeenth century scientific revolution, many of whom were committed Christians. John Wilkins, a founder member of the Royal Society, wrote over 300 years ago³: "It were happy for us, if we could exempt Scripture from Philosophical controversies: If we could be content to let it be perfect for that end unto which it was intended, for a rule of our Faith and Obedience, and not stretch it also to be a Judge of such Natural Truths as are to be found out by our own Industry and Experience".

Similar components may be found in the writings of Kepler, Galileo and Boyle, and indeed this view has been the mainstream Christian view down the centuries.

DENIS R. ALEXANDER

*Imperial Cancer Research Fund
Laboratories,
PO Box No. 123,
Lincoln's Inn Fields,
London WC2A 3PX, UK*

1. Chadwick, O. in *The Victorian Church*, Part 2, 23-35 (Black, London).
2. Livingstone, D. N. *Darwin's Forgotten Defenders* (Eerdmans/Scottish Academic, Edinburgh, 1987).
3. Hooykaas, R. *Religion and the Rise of Modern Science*. 116 (Scottish Academic, Edinburgh, 1972).

Dual-role Hanford

SIR—The leading article "Nuclear togetherness" (*Nature* 334, 551; 1988) seems to ignore the fact that the Hanford nuclear reactor has, for more than 20 years, produced both plutonium for military purposes and electricity for the Washington public power supply system. Thus the US Department of Energy's proposal to make the modular high temperature gas-cooled reactor (HTGR) dual-purpose is entirely consistent with well-established practice.

The modular HTGR achieves a level of passive safety that far exceeds that of any existing commercial reactor. As such it offers an opportunity for a new era of nuclear energy based on reactors for which the possibility of a Three-Mile Island, let alone a Chernobyl, is practically nil.

The Department of Energy, and Senator McClure, rather than being chided, should be congratulated for having proposed a practical path for developing this super-safe reactor technology at a time when commercial nuclear energy development is almost at a standstill in the United States.

ALVIN M. WEINBERG

*Institute for Energy Analysis,
Oak Ridge Associated Universities,
PO Box 17, Oak Ridge,
Tennessee 37831-0117, USA*

Price of freedom

SIR—The leading article by John Maddox, "What price academic freedom?" (*Nature* 334, 377; 1988) is a timely eye-opener to academics anywhere. The ultimate price is of course dismissal. Many of us at the University of Malta were in fact dismissed in 1978. It is not quite correct to say that the university gave in to the government of the day. The comparison of academics with farmers does not do justice to farmers. Firmness, as any farmer can tell academics, can stop academic rot.

W. H. BANNISTER

*Department of Biomedical Sciences,
University of Malta,
Msida, Malta*

Nothing signifying

SIR—Following the recent correspondence in *Nature*, I get the impression, that there must be something about homeopathy. I don't believe that no-more-existent molecules can leave an imprint in water. But isn't it striking, how a non-existent phenomenon can leave an imprint in the scientific discussion?

KARL O. ROTHHAUPT

*Max-Planck-Institut für Limnologie,
Abteilung Ökophysiologie,
Postfach 165,
D-2320 Plön, FRG*

Benveniste on the Benveniste affair

Dr Jacques Benveniste replies to some of the points raised in recent correspondence in a text which is printed (as on a previous occasion) unchanged except for grammatical reasons.

Paris

WHAT a deluge! Certainly appropriate about a paper on water but a lot of it could have been deleted using *Nature's* criteria for publication. Our publication of our paper was a cry for help to explain these puzzling results. Instead we got a fraud squad, an unsound (a fraud implicating five laboratories?) insult to respected scientists, who from the start and thereafter were treated as criminals. All scientists on Earth should resent the insult or at least the threat.

The style of the report, some of the expressions used, the sensational titles — indeed, it's a shame. A well-done job, admittedly, that condemns us first, then covers pages that discouraged many readers. And, shamelessly, a critical sentence indicating that many (?) of our results are statistically correct was removed at the last minute, after receiving my answer (*Nature* 334, 291, column 3, paragraph 2).

A section called "Collaboration" was also added at the last minute which is filled with "mistruths": data from Israel, twice described as not available, can be found . . . in our *Nature* paper (Table 2), and the corresponding raw data were given to *Nature* editors in March 1987. The report by Maitre Simart, the bailiff, describing the coding and decoding in April, May and June 1987, in the presence of the dean of our faculty, is available to anyone.

Since, to our surprise, the unbelievable *Nature* report was taken by some as evidence negating our data, we have to repeat once more the litany of the amateurish mistakes of the "experts" that "do not match their extraordinary claim" of being the world scientific conscience and judge: confusion of single-code done twice with double-blind, with circus-like — and fraud-seeking — pantomime of sticking the tape-armoured code to the ceiling (why not in their pocket since they knew the code? Because they wanted to catch the villains tampering with the sacred paper, hence the "expert at opening envelopes"), participation of the referee in the ball-game (that he had never played before), misquotation in the report of their own figures, erroneous Fig. 6 captions and the immortal "we have a record of the proceedings on an *unbroken* reel of tape".

Yet the most salient feature of this report is confirmation of our data by two positive experiments (Fig. 2 and Fig. 6)

that remain unchallenged. I am supposed not to have ever seen anything such as Fig. 2 before, when the same is in Table 1 and Fig. 1b of our paper. The two plots in Fig. 6 are declared discordant, exactly what is printed in our article.

The following correspondence for weeks occupied precious *Nature* space to show both that these data do not exist and how to explain them. I am more concerned at Gaylarde's claim that they are synthetic (334, 375), which is entirely unsubstantiated as well as being a blow to academic courtesy.

It is clear from the Table 1 legend of our paper that dilutions were double-coded but not the reading. This could explain an unconscious gap-closing between triplicates but not the differences between coded dilutions and controls. The fourth experiment (Fig. 2 of the report), read blind, yielded "unbelievable", "incredible" replicates that cannot be found in the report. Gaylarde's next argument is indeed damning, but to whom? Counts higher than controls were not ignored but, as usual when using percentages, were equated to 100 since it does not make sense to create basophils. We made clear in our paper that statistics were done on actual numbers and not on percentages.

Metzger (*ibid.*) and Seagrave (334, 559) show no numbers over 100, but are immune from criticism. In the latter letters, and that of Bonini (334, 559), small numbers of experiments are enough to show we are wrong. But our exact experimental design was not reproduced. Cell biologists know that release of mediator is a late event preceded by myriads of intracellular signals. Our test monitors at the granule matrix level cation competition with a positively charged dye. In fact, we have occasionally also seen histamine release but this is hardly reproducible. Using mediator release to confirm our experiment is as sound as reproducing data on muscle ATP by studying contraction.

Several letters propose hypotheses wilder than our data: the antibody-mimicking heparin, a mysterious degranulating molecule that yields a rhythmic fluctuation that is not so rhythmic, cavitation, free radicals. . . . Answer to these entertaining fantasies is in Fig. 1b of our paper.

To try to raise the tone of the debate, I will take the goat/unicorn story from Randi and the request of "different edito-

rial standards" for extraordinary data from H. Metzger. J. Randi believes that it takes special data to tell a unicorn from a goat. In the experimental process, data must be collected and interpreted independently of the weight and implication of the hypothesis. A goat is a unicorn when a statistically significant number of experiments have shown a unique horn, obviously after Randi has checked that it was not glued with Araldite. If there are two horns, it's a goat, regardless of electron microscopy, NMR imaging or gene splicing.

This is serious matter. Changing the rules of experimental science will first kill fragile data critical to fringe advances and then science as a whole. Let us not depart from this "fundamental principle of scientific objectivity". Finally, some good news. The riddle is solved: somebody came in our backyard and shot the unicorn.

Indeed, only the smile is left . . . with a question: why? It seems illogical that having scrutinized the paper for two years, having urged confirmation of our initial work in independent laboratories, which was done in Canada, Israel and Italy. *Nature*, short of any valid objection, published it hastily, to then go these extreme lengths to . . . de*Nature* it. The answer is to be found both in *Nature* 333, 787, para. 4, and in the report (334, 287, para. 5), both of which emphasize warnings against homoeopathy.

Fact twisting, errors, omissions, misquotations and mistruths are symptoms of a crusade. Also revealing is that at the question "why publish before the inquiry", the *Nature* staff tells interviewers: "Benveniste would have withdrawn the paper". Clearly the plot was to scare the bird out of the bush to shoot at ease. Their indiscriminate fire is typical of their disarray in front of these obviously positive results.

John Maddox (*New York Times* 28 September) still declares them non-existent, whereas J. Randi now "admits" (*Espresso*, Lisbon, 1 October) that it is in fact a fraud. A fraud with five laboratories and no results! What on Earth is this mess? Unfortunately, facts are stubborn and so are we. The numerous truth-seeking scientists all over the world, some of them prompted by our paper and the obviously biased inquiry, have intellectual and technical means either to understand the error or to establish this new field. There is more to come.

Jacques Benveniste

Waves caused by extreme dilution

From this issue of Nature, the past several weeks' correspondence on the Benveniste affair will be closed. The editor of Nature here discusses some of the issues that have arisen.

IN FIFTEEN out of the past twenty-two years as editor of *Nature*, I have known nothing like the controversy touched off by the publication last June of an article by Dr Jacques Benveniste and a group of his associates (Davenas, E. *et al.* *Nature* **333**, 816; 1988) claiming that solutions of anti-IgE retain their biological effectiveness even when indefinitely diluted, and by the publication, a month later, of the report of an on-site investigation by three people, myself among them, who have no expertise in the field of allergic immunology (*Nature* **334**, 287; 1988). I have learned a great deal from the controversy, as have many of my colleagues, although I do not pretend fully to understand why such great passions have been aroused.

What follows is a discussion of some of the issues that have arisen. Because it is intended that *Nature's* correspondence columns should now revert to their normal uses, only matters raised in the original papers, the subsequent correspondence and certain newspaper comments have been dealt with. New material has been avoided. It is hoped that readers will nevertheless find this account an explanation of some issues not previously fully dealt with. It may be found also to point to some questions about the functioning of the scientific literature and even the practice of science for which there are no simple answers.

The essence of Benveniste's claims is easily summarized. The immunoglobulin IgE is one of the mediators of allergic reactions in mammals, at least in part through its attachment to the surfaces of mast cells and of the subclass of lymphocytes known as basophils. Allergens, for example pollen grains, in the blood of persons carrying IgE of appropriate specificity will react in such a way as to trigger first the expulsion of (exocytosis), and then the release of histamine and other humoral factors from, the vesicles containing them in both mast cells and basophils. The same effect can be brought about by antibodies against IgE, on at least one view because they induce the formation of a network of crosslinked IgE molecules on the cell surface.

Anti-IgE antibody is prepared by injecting human IgE into some other mammalian species; Benveniste's material, manufactured by the Dutch company Nordic Immunology, is prepared in goats. The antibody can be used as a reagent in the assessment of allergic

reactions in human patients, for example, in whom the gross amount of circulating IgE may be an indicator of a patient's allergic status.

Experiments

The standard measurements at Benveniste's laboratory, INSERM 200 at Clamart in Paris, as observed in July and recorded in earlier laboratory protocols, involves the interaction of anti-IgE solutions with samples of white blood cells taken from the blood of human volunteers. Intact basophils are known to be stained red by their interaction with an acidified solution of toluidine blue. The

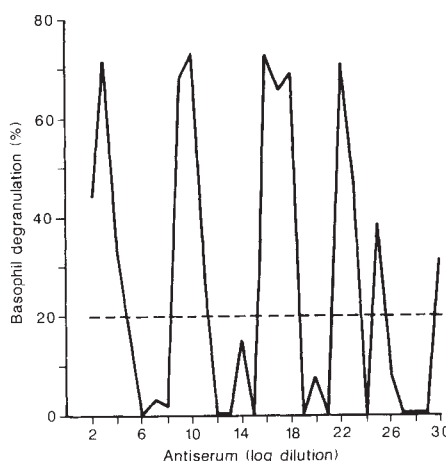


Fig. 1 The fourth of seven demonstration experiments (read 'blind') with unexpectedly high peaks.

standard experiment consists first of allowing suspensions of white blood cells in buffer (a version of Tyrode's solution) to interact with serial dilutions of IgE in the same material and then of counting the surviving intact basophils after staining with toluidine blue.

The general conclusions will be well-remembered. In experiments in which anti-IgE solutions undergo a sequence of ten-fold dilutions (simply accomplished by measuring a fixed amount of inert solvent into each of a sequence of test-tubes and then transferring one ninth of the contents of one tube into the next in line), some of the diluted solutions are found to retain the effectiveness of undiluted IgE in degranulating basophils. The experiments we witnessed in Paris stopped short at 35 successive dilutions, but the original paper referred to experiments in which 60 successive ten-fold dilutions had been carried out, and in

which dilutions of 10^{130} had been reached by successive 100-fold dilutions.

The significance of these numbers may not be fully appreciated (although it was clearly explained in the original paper). Avogadro's number is 6.02×10^{23} and the concentration of the commercially supplied anti-IgE is 1 mg cm^{-3} , which corresponds to a molar concentration of $2.2 \times 10^{-6} \text{ M}$. One cm^3 of the commercial preparation will therefore contain some 1.5×10^{15} molecules of anti-IgE.

Provided that the number of antibody molecules transferred from one tube to the next is proportional only to their concentration in the first tube and the volume of the liquid transferred, after 15 ten-fold dilutions, 1 cm^3 of solution will on the average contain 1.5 molecules of anti-IgE — the actual number will be a matter of chance governed, no doubt, by a Poisson distribution. Further dilution beyond this point will yield solutions in which the average number of antibody molecules in 1 cm^3 will be less than one.

It is not for nothing that one correspondent has raised the question of how these extreme dilutions square with the supposedly finite size of the Universe. If the cosmological view that there are 10^{77} baryons (protons and neutrons) in the Universe is correct, the whole bulk of the Universe even if converted into water would be insufficient to dilute the contents of a 1 cm^3 vial of anti-IgE to the highest dilutions of 10^{130} reported by Benveniste.

The second striking feature of the original paper is the assertion of a rhythmic appearance and disappearance of the degranulation effect at successive dilutions. In the original paper, it was reported that "the repetitive waves of anti-IgE-induced degranulation were reproducible, but the peaks of degranulation could shift by one or two dilutions with every fresh sequential dilution of anti-IgE and depended on the blood sample". The phenomenon is said to depend crucially on the agitation of the freshly diluted solutions at each stage.

No firm explanation of the phenomenon was put forward, but it was suggested that "water could act as a 'template' for the molecule . . .", in effect a kind of memory for molecules long-since diluted away. No model to account for the rhythmic fluctuations of degranulation has yet been put forward. As a reminder of the data, Fig. 1 shows a typical plot of basophil degranulation against anti-IgE dilu-

tion — in reality, this represents the data the fourth of seven of the sets of measurements gathered during July.

While others working in the field say that it would have been better to have worked with mast cells than with basophils, or that basophil degranulation can be more reliably estimated by measuring histamine release quantitatively than by counting stained cells, it should be recalled that the INSERM 200 group acknowledged in July that it had been found that the disappearance of stained cells after reaction with anti-IgE at extreme dilution does not correlate well with histamine release, and cannot therefore be used as a measure of phenomenon.

Nothing in this account of the standard experiments would give offence at INSERM 200.

General criticisms

At the outset, the torrent of correspondence generated by these publications consisted mostly of the responses to the invitation to readers to suggest how the data reported might have arisen normally, perhaps as artefacts, but from the start there were also some general criticisms of *Nature's* handling of the affair. Some of the questions raised are summarized below, together with my responses. Because many of the questions raised by correspondents and others are contingent on each other (as in "If you had to publish, why not investigate first?"), this dissection of the problem may over-simplify the issues that most trouble readers. It will become apparent that the responses hang to some degree on a conception of *Nature's* function of which even regular readers may not be fully aware.

Why publish at all? or, in a stronger version, *Why publish with reservations so explicit as to suggest the conclusions should not be believed?* All journals know that, while most articles submitted for publication can be dealt with quickly, there are some on which referees cannot offer helpful advice. Over a period of nearly two years, there were in this case four referees, of whom three provided detailed reports on various occasions. The general tenor of all comments was that, if the data were correct, INSERM 200 must indeed have discovered a remarkable phenomenon, but that the conclusions were so remarkable that the experiments must in some way be flawed.

Throughout this process, Dr Jacques Benveniste (with whom all correspondence was conducted) appeared readily to follow all suggestions offered, in particular arranging that measurements should be independently repeated; he chose laboratories in Israel and Milan. Copies of pages from the Clamart laboratory notebooks as well as data from Israel and Milan were sent for scrutiny by referees.

In retrospect, it is nevertheless clear

that the paper as published contained internal inconsistencies, notably that quoted errors were much smaller than the expected sampling errors. More seriously, perhaps, we were unimaginative with our suggestion of how more stringent controls might have been devised. I should also have been more cautious when, having rejected the paper for what my colleagues hoped would be the last time, Dr Benveniste telephoned indignantly to protest that *Nature* was proposing to suppress news of one of the greatest discoveries of the twentieth century. I forget whether he compared his dilemma to that of Galileo on that occasion or in a conversation during the later visit to Paris.

To my complaint along the lines of "But you don't even consider how these extraordinary results might be conventionally explained" Dr Benveniste answered with a version of the Russian verbal shoulder-shrug "No problem!"; the section of the published paper about the long-term memory of water followed only a few days later.

Journals are these days painfully aware of the accusation that their collective influence is to inhibit innovation. The accusation is untrue, given the competition among journals for soundly based innovative publications. *Nature*, moreover, consciously publishes a regular sprinkling of heterodoxy, mostly as "Commentary" or "Scientific Correspondence". This is an important function even when there is little chance that the heterodox will become the orthodox; people may find it instructive to know what is happening on the fringes of their interests.

But the accusation of suppression by the head of a government-supported laboratory cannot be dismissed lightly. And while, in an ideal world, exasperation should play no part in editorial decisions, it must surely on occasion be proper that a person convinced that heterodox conclusions are correct should be given the opportunity for which he asks to see what happens when they are tested publicly.

Those were the circumstances in which it was agreed between Dr Benveniste and I that the paper would be published, but that publication would be followed by an on-site investigation by a nominated group. From the outset, Dr Benveniste knew the composition of the group and that it would be sceptical of his conclusions. My concern was that the publication of his paper (certain to excite the interest of the homoeopathic community) should be followed as quickly as possible by the appearance of a report of the investigation. The actual timing of the visit to Paris was decided by Dr Benveniste and Mr James Randi; for *Nature's* purposes, it was inconveniently soon.

Why not investigate first and publish only afterwards? Journals do not normally

undertake investigations of contributors' laboratories, and for good reason: they have neither the resources nor the skill, and they cannot command them from elsewhere. The decision to take up Dr Benveniste's long-standing invitation to "send" a group of people to Paris was mine, prompted largely by the exceptional circumstances — Benveniste's evident conviction of the correctness of his conclusions, his insistence that his data should see the light of day and our suspicion that one explanation might be that the data had been generated by a hoaxer in his laboratory. It seemed best that the material should be published before and not after the investigation. With hindsight, it would probably have been sufficient that there should have been an informal preliminary visit to the laboratory, but that was not self-evident at the outset.

From correspondence and comment in the general press, the procedure that we actually followed seems to have raised objections on two counts. First, it is said, it has turned a serious issue into a "circus". Second, the procedure was essentially a trick by means of which Dr Benveniste and his colleagues exposed themselves to gratuitous criticism (or were "stitched up", as the *New Scientist* inelegantly put that point in August).

Bargain

In reality, the arrangement was an explicit bargain. Investigation was a pre-condition of publication (although Dr Benveniste appears not to have informed his colleagues of that circumstance). But even though avowedly sceptical of the conclusions, I had no reason for confidence that a necessarily brief visit to Paris would have yielded support for that scepticism. That (in our opinion) it did has inevitably engendered something of an air of melodrama (the extent of which is nevertheless surprising). But those who have advanced this criticism overlook the danger nearer the front of my mind in June — the chance that a brief visit to INSERM 200 would provide substance only for a report recording that experiments were carried out, and with the results already anticipated in *Nature*.

To have inverted the bargain would have given Dr Benveniste full freedom of decision; given a favourable or even non-committal report, he would have claimed publication, but otherwise could have withdrawn. Locking both of us into an agreement to publish simultaneously would have led either to an implicit endorsement of an unbelievable result or to a nonsense — the simultaneous publication of an extraordinary claim and its purported refutation. (The circumstances are different from those of the earlier simultaneous publication of such a claim, that a protein called scotophobin can serve as a vehicle for the transfer of learned be-

haviour from one rat to another, and its refutation, in that the refutation sprang from the internal inconsistencies of the documents submitted for publication.) Even as things were arranged, the investigation might have found no obvious fault with the conduct of the experiments — an outcome of which Dr Benveniste appeared confident until almost the end.

Nature has no ambition to lead a pack of vigilantes seeking to rid the scientific literature of error, and worse; it is rather concerned that over-zealousness may yet make honest mistakes culpable. But there is also a distinction to be drawn between erroneous science, however motivated, and science that is conducted carelessly, allowing sharp inferences to be drawn from insubstantial data.

I am puzzled that Dr Benveniste is as indifferent as appears to be the case, both in several conversations in Paris and in his two comments on our report, of the complaint that he and his colleagues were unaware of the importance of sampling errors. At our final conversation on 8 July, it was clear that the relevance of the point was simply not understood, and discounted as a "theoretical objection".

To the extent that allowance for sampling errors would have made some apparently statistically significant measurements statistically insignificant, it may seem proper simply to set those disputed data aside, relying on the remainder to sustain the startling hypothesis. This is what Dr Benveniste does in referring to "confirmation of our data by two positive experiments", where the Fig. 2 referred to is reproduced here as Fig. 1.

But what if indifference to an ubiquitous source of error should lead to the uncritical acceptance of data which appear to be more consistent among themselves than the simple arithmetic of sampling error would allow? Fig. 4 in our report (reproduced here as Fig. 2) is compiled from all multiple measurements of the same samples recorded in the notebooks. Its striking feature is that the distribution of the discrepancies of measurement is, for whatever reason, narrower than the gaussian distribution expected for sampling errors. The bite chewed out of the measured curve at its mid-point even suggests that identical measurements of samples of the same suspension of blood cells are less frequent than sampling theory would allow (and dictate). These distributions suggest that observer bias seriously affects the whole series of data: in the circumstances, it is difficult to tell what confidence there can be in even the most apparently clear-cut graph, even the data represented here in Fig. 1.

Figure 1, on the other hand, has been taken by Dr Benveniste in his accompanying comment and elsewhere as a proof that even a hostile investigation threw up data consistent with his hypothesis and he

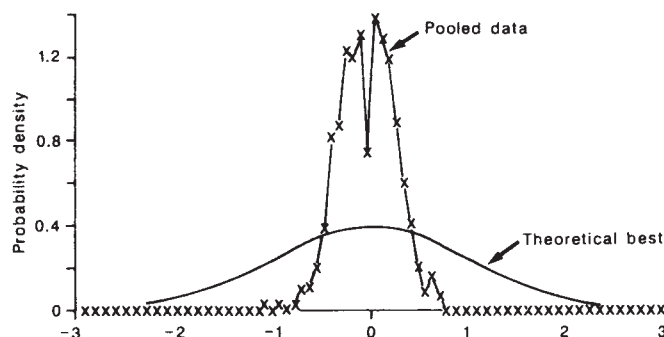


Fig. 2 Comparison of measured departures of duplicate normalized readings from their means with the gaussian distribution expected (Fig. 4 of the report by J. Maddox *et al.*)

denies (contrary to the recollection of all three of us) that he remarked "We've never seen one like that before". But inspection shows that there is a significant difference: in Fig. 1, the data points are either very low (no degranulation) or near the maximum (70 per cent or more degranulation). In other published curves and in the laboratory notebooks, the much more common pattern is that there are intermediate data-points, lying impossibly often on straight lines between a peak and a trough.

In the scientific literature, outright fraud is rare, honest error is probably much more common but is commonly corrected either by its authors (or, more eagerly, by their competitors) or, more often still, is ignored, to the general confusion. But the fate of careless science is usually more seemly — it earns its authors a modest degree of acclaim, but is in the end usually forgotten. Dr Benveniste's data have been used to sustain an interpretation that cannot easily be overlooked, but they are also in our judgement carelessly gathered and carelessly interpreted. In my opinion, if these complaints are true, it is a matter of some importance for the scientific community that the extent of these lapses from good practice should be more widely known.

Conviction

There is every reason to believe that Dr Benveniste is convinced of the reality of the phenomenon he describes, and that he is convinced that the data have been gathered properly. But the essence of the report of our investigation is that the Clamart team did not appreciate that sampling errors are unavoidable in their measurements, had not investigated the causes of qualitative variability in their experiments and worked in an intellectual climate conducive to observer bias. Dr Benveniste appears not to have dealt with these charges either in his two published statements or in statements quoted in the general press.

Why send a team of amateurs? Although the composition of the visiting group was known to INSERM 200 at the outset, a fuller explanation is called for, especially because the composition of the group is

another of the reasons why the visit to Paris has been called a "circus". That the composition of the group was unorthodox was acknowledged in July. Even so, its members' credentials are not inconsiderable.

The short answer to the question is that if a group of mere amateurs can so quickly discover procedural errors of such importance, that is sufficient justification.

More fully, Walter W. Stewart was one of the most assiduous of the referees of the Benveniste manuscript in its various forms. Over the years, he has demonstrated a flair for spotting inconsistencies in intricate arguments. Stewart has demonstrated, in his handful of published works, great skill as a laboratory scientist. His friends (and his employers, the National Institutes of Health, no doubt) would wish that the zeal he has shown in the pursuit of error in the literature had been devoted to some substantial scientific project. Stewart's presence in Paris has been understood by some in the United States to be an endorsement by *Nature* of his role in the recent pursuit of Dr David Baltimore for not having disowned an alleged error in a published paper (see *Nature* 333, 795; 1988), which is mistaken.

James Randi, whose presence seems most to have offended Dr Benveniste, is more than a mere stage performer. His role in demonstrating that Uri Geller's illusions could be accomplished by conjuring is now an acknowledged public service, since when he has devoted a large part of his considerable energy to exposing other exploitations of public credulity. That he is a MacArthur Fellow is well-known, that his first job was as a laboratory assistant at the Banting and Best laboratory in Toronto is less so. His presence in Paris, originally suggested as a means of telling whether Dr Benveniste was being hoaxed, proved much more valuable than that. Over the years, Randi has developed a talent for planning in advance how to carry out procedures and to prove that they have been followed faithfully. Even though he announced on 5 July that his task was over, he applied his great intelligence to every detail of the procedures carried out at Clamart. It was he, for example, who warned us that if

Stewart were to carry out the pipetting of blood samples (one of the obvious ways in which readings might be corrupted by accident or design), it was essential that one of Dr Benveniste's colleagues should keep him closely under watch. Unfortunately, Dr Benveniste appeared not to appreciate that Randi is more than a mere conjuror.

Negotiations

Incredible though it may seem, my own intention has been to spend only a day or two in Paris at the beginning of the investigation, helping to negotiate the programme of work for the days ahead and reaching an understanding with everybody concerned about such matters as communicating with the press (an undertaking that Dr Benveniste, unfortunately, was unable to keep). But it quickly became apparent that the process of negotiation would have to be continuous.

It goes without saying that the group's report purports to provide only a sufficient reason why the data cannot be held to sustain the conclusions built on them, and not an explanation of those that appear to stand out above the noise. It may well be that there is also some kind of artefact buried in the data, but if the significance of the data is consistently overestimated by the neglect of sampling errors, a search for artefacts will require the recompilation of a comparable dataset under more rigorous conditions.

But in one important respect, our investigation and report were deficient: we were unconstructive in our comments. On reflection, it is plain that there are several important sources of variability in the measurements that require fuller investigation. Thus there seemed to be a curious lack of intellectual curiosity at Clamart about the reasons why some bloods do not degranulate on some days, why some violent agitation — "vortexing" — is necessary after each dilution, why the results of experiments are better after blood cells have stood in the cold room overnight, and so on. All these would seem crucial both to understanding and underpinning the physical conclusion that extreme dilution has no effect, as would be the further investigation of whether the distribution of anti-IgE is homogenous in the dilution procedure followed in the experiments (one such experiment, with radio-labelled antibody, is referred to in Davenas *et al.*).

What of the other laboratories? Our neglect of what Dr Benveniste considers to be confirmatory evidence appears to have caused him the most distress, and for reasons which are readily understood. Three (not five) collaborating groups were referred to in the original paper.

The data available from the Israeli work is the most explicit but also somewhat confusing. We know of three separate phases of investigation — an attempt to repeat

the Clamart experiments (with negative results), a further trial in the presence of Elisabeth Davenas (which yielded positive results but also, unfortunately, accusations of deception by some members of the Israeli group) and a further trial organised remotely from Paris under the supervision of the Clamart bailiff, M. Simart.

The data from the second trial are undoubtedly significant; we said so. There is a profound misunderstanding about the third series of measurements, whose incompleteness came to light when we failed to find the decoded data in the notebooks we had borrowed. Our recollection is that Dr Davenas said at our meeting on 8 July that M. Simart had been too busy to decode them, and that Dr Benveniste said something to the effect that "I'll get them from him on Monday". But now, members of the Paris and the Israeli groups have said that the data were already decoded, in which case we have not seen them (or have mistaken them for other data).

Difficulties

In the hope of circumventing these difficulties, I wrote to all Dr Benveniste's co-authors some weeks ago, asking whether they wished to comment on the past few months' events, for publication or otherwise. One member of the Israeli group has replied, presumably on behalf of all, but unfortunately has stipulated that the contents of his letter must be confidential. The Milan group says it considers *Nature's* handling of its contributions "unfair" and, in passing, that it had not been aware of the referees' criticisms until the publication of Davenas *et al.* in June. The Toronto group (whose work was cited as confirmatory) replied with a series of legitimate questions about our report, to which there has not been time to provide replies; but the group appears to be embarking on an objective study of the phenomenon in what one member described as a "Popperian spirit".

Nature will be glad to publish as Scientific Correspondence the general conclusions of any or all of these groups when they are ready.

Has it been worthwhile? The short answer is "No". Too many people have been caused too much distress, particularly at INSERM 200, and too much time has been consumed by the circus when there have been better things to do. But again the longer answer is that there are several ways in which the affair has been instructive, certainly for those involved and, it must be hoped, for many readers.

My own guess is that Dr Benveniste's colleagues will now be counting basophils in replicate, following the standard procedure for controlling sampling errors, and will be eliminating unavoidable observer bias by making blind measurements a routine. I expect that the results will not

differ substantially from those obtained in the three blind experiments (each with two observers) at Clamart on 9 and 10 July; it will be extremely interesting if it should be otherwise, but no doubt Dr Benveniste would prefer to publish that intelligence in some other journal.

Fallibility

Nature has learned much about the fallibility of the refereeing process, which is not to suggest that the referees involved in this extreme dilution case have been anything but excellent. To put the matter uncontentiously, there are circumstances in which it is difficult to tell from the inspection of a manuscript, and of proffered supplementary material, whether some kinds of claims are valid. This does not imply that the refereeing process is uniformly unreliable, or even consistently full of holes, but that there must be circumstances in which the decision whether or not to publish is subjective. There is no surprise in that for those with editorial responsibility, but — even if only marginally — it gives the lie to the myth that what appears in the journals is the next best thing to the truth.

One of the uncomfortable findings of the high-dilution matter is nevertheless that *Nature* appears to be shackled in the public mind to the totem-pole of absolute veracity. That seems partly to explain much of the reaction of the general press to the events of the past few months. Its interest has been intense but superficial. On one occasion I found myself being questioned closely by a reporter using an *aide-memoire* supplied by Dr Benveniste. It is sad that the notion that truth is black or white takes so long to disappear.

Critics from among the research community have had different motives. Publication has become one of the chief means by which standards of quality are set and maintained in the research profession. *Nature*, with no formal connection with academies and professional societies, has nevertheless by tradition acquired a licence to be a part of the process, to the general enlightenment of its readers. There is a sense in which part of the price of the independence is to risk innovation, which explains why *Nature* sometimes appears to behave brashly. But what has plainly seemed to some to have been a mockery of the standards-setting process, however instructive in other ways, is bound to have engendered unusual passion. Thus it is that the events of the past few months are unlikely to be repeated regularly.

So what is the truth about INSERM 200's claims on behalf of high-dilution anti-IgE? One correspondent chided us with having impeded the discovery of the true explanation. My own conviction is that it remains to be shown that there is a phenomenon to be explained.

John Maddox

Millisecond pulsars

Shock wave reveals pulsar wind

J. C. Raymond

OLD pulsars with millisecond periods are remarkable end-products of binary-star evolution: neutron stars which have been spun up by accretion. The initial rotational energy of a millisecond pulsar is as great as the kinetic energy of the supernova which originally created the neutron star, and the observed rates of period increase imply that this rotational energy is lost at a rate greater than 10^{35} erg s^{-1} , probably as a wind of relativistic particles. These winds are surprisingly difficult to detect. Soft X-ray nebulae have been found around a few pulsars¹, but most could not be seen. Kulkarni and Hester report on page 801 of this issue² the detection of another manifestation of a pulsar wind, a faint nebula of pure hydrogen Balmer-line ($H\alpha$) emission, marking a shock wave that the wind drives into cold interstellar gas.

Hester and Kulkarni have previously studied³ the emission nebula created when a rapidly moving pulsar strikes the dense shell of its own supernova remnant, CTB 80. The new nebula encloses the recently discovered eclipsing binary pulsar PSR1957+20. The orbital period and velocity amplitude of this object imply that the binary partner has an extremely low mass (about 0.022 solar masses), yet the eclipse duration requires it to be larger than the Sun and opaque to radio waves. This can be understood if the opaque object is actually a wind evaporating from the low-mass companion⁴.

Energetic electrons and positrons in the pulsar wind can produce enough γ -ray emission in a shock around the secondary to heat its surface and drive the wind. The evaporation may be rapid enough to destroy the secondary entirely, explaining the existence of solitary millisecond pulsars^{5,6} and the dearth of short-period low-mass X-ray binaries⁷.

The $H\alpha$ nebula around PSR1957+20 reported in this issue² is another consequence of the pulsar wind. It is very faint and the quality of the image (see page 802) demonstrates the sensitivity of the current generation of instruments. Its bow-shock shape results from the motion of the pulsar, at a velocity v of about 100 km s^{-1} , through the ambient gas. The distance between the pulsar and the bow shock is given by the balance of the ram pressure, ρv^2 , of the bow shock (ρ is the density of the interstellar gas) and the pressure of the pulsar wind. The pure $H\alpha$ emission spectrum is interpreted as a non-radiative shock (a shock for which radiative cooling is not dynamically important), similar to those observed in a few young

supernova remnants. The Balmer-line emission is produced by collisional excitation of neutral hydrogen swept up by the shock, and if the line profile can be measured, its width determines the shock velocity⁸.

This interpretation imposes very strict limits on the pulsar-wind emission and on the nature of the ambient interstellar gas. The presence of neutral hydrogen so close to the pulsar strongly suggests there is a low-energy cutoff to the power-law emission of the wind. Some difficulties may arise in applying the existing models of non-radiative shocks. The intensity ratio of the two Balmer lines $H\alpha/H\beta$ is much higher than in model predictions, and even the non-radiative nature of the

shock is barely consistent with the size and density of the nebula.

Although these difficulties could indicate an interaction between the pulsar-wind particles and the interstellar gas more direct than a pressure-driven shock wave, the most exciting aspect of this discovery is the opportunity to measure the impact of a pulsar wind on the interstellar medium. □

1. Helfand, D.J. in *Supernova Remnants and their X-ray Emission* (eds Danziger, J. & Gorenstein, P.) 471-484 (Reidel, Dordrecht, 1983).
2. Kulkarni, S.R. & Hester, J.J. *Nature* **335**, 801-803 (1988).
3. Hester, J.J. & Kulkarni, S.R. *Astrophys. J.* **331**, L121-L125 (1988).
4. Fruchter, A.S., Stinebring, D.R. & Taylor, J.H. *Nature* **333**, 237-239 (1988).
5. Phinney, E.S., Evans, C.R., Blandford, R.D. & Kulkarni, S.R. *Nature* **333**, 832-834 (1988).
6. Kluzniak, W., Ruderman, M., Shaham, J. & Tavani, M. *Nature* **334**, 225-227 (1988).
7. van den Heuvel, E.P.J. & van Paradijs, J. *Nature* **334**, 227-228 (1988).
8. Chevalier, R.A. & Raymond, J.C. *Astrophys. J.* **225**, L27-L30 (1978).

J. C. Raymond is at the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts, 02138, USA.

IN BIRDS, the most common clutch size varies from one species to the next; gannets, for instance, tend to incubate eggs singly, whereas great tits typically incubate half a dozen or more eggs. Moreover, there is considerable variation in clutch size even within some species. In Wytham Wood near Oxford, for example, the actual number of eggs laid by great tits varies from five to a dozen. How is this inter- and intra-specific variation to be explained? Group selectionists such as V.C. Wynne Edwards have argued that clutch size is regulated so as 'altruistically' to avoid population overcrowding. Followers of David Lack, however, argue that clutch size is adjusted by natural selection so as

to maximize the number of surviving offspring. The picture shows a male collared flycatcher (*Ficedula albicollis*): on page 813 of this issue, Lars Gustafsson and William Sutherland report the results of experimental manipulations of actual clutch size in this species, and conclude that large clutches have considerable concomitant costs. In a separate study soon to be published in *Nature*, R.A. Pettifor, C.M. Perrins and R.H. McCleery report that individual female great tits (*Parus major*) lay that size of clutch from which they can maximize the number of recruits (young which survive to enter the subsequent breeding population). Despite the differences evident in these two studies, together they suggest that clutch size has evolved by individual rather than by group selection. (Painting by Lars Jonsson.)

Rory Howlett



Cervical cancer

News on papillomaviruses

D. J. McCance

Most infections caused by human papillomaviruses, which often result in the excessive epithelial proliferations called warts, are, like beauty, only skin deep. There are more than 50 types of these viruses, and it is becoming increasingly clear that some types, those producing genital lesions and transmitted sexually, are the aetiological agents of cervical cancer. Various environmental factors have been implicated in this relationship, for example, the report by Pater *et al.* on page 832 of this issue¹ suggests there is a link between cervical cancer and glucocorticoids.

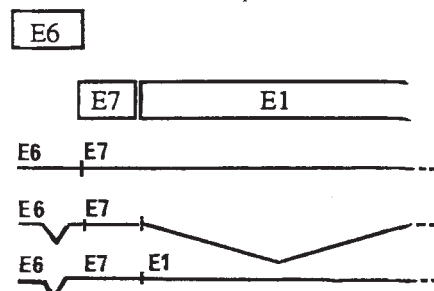
The idea that sexually transmitted infectious agent(s) cause cervical cancer stems from the fact that the greatest single risk factor for the disease is the number of sexual partners a female or her single male partner have encountered. Cervical cancer arises in the transformation zone of the cervix, an area of immature metaplasia between the mature squamous epithelium of the exocervix and the columnar epithelium of the endocervix canal. The premalignant lesion, termed cervical intraepithelial neoplasia (CIN) is characterized by a lack of differentiation of stratified squamous epithelial cells and is graded from I to II in increasing severity, with CIN III exhibiting little or no differentiation.

Although genital human papillomavirus (HPV) types 6 and 11 are associated with benign genital warts and low grades of CIN, HPV16 and 18 are associated with most CIN lesions of all grades and invasive cancer. The viral DNA detected in premalignant CIN lesions is episomal, freely replicating and produces infectious virus, whereas in malignant cells the DNA is randomly integrated into various chromosomes. It appears, therefore, that there are two groups of HPV — one with low and one with high oncogenic potential.

Although it is becoming increasingly accepted that some HPV types are involved in the development of the premalignant disease, other factors may be necessary for progression to malignancy. One independent risk factor is smoking and another less obvious one is the oral contraceptive pill, although the epidemiological evidence for the latter is contentious. The report by Pater *et al.* in this issue¹ suggests that addition of glucocorticosteroids to culture medium increases the transforming ability of HPV16 (high-risk type) but not HPV11 (low-risk type) when under their endogenous promoters and in cooperation with an activated *ras* oncogene. The authors interpret

their results to suggest urgent studies on the effect of oral contraceptives on the development of cervical cancer.

It has previously been shown that the non-coding region of HPV16 possesses an area containing promoter and enhancer elements, a glucocorticosteroid-reactive element (GRE), which is activated by dexamethasone with resulting increase in transcription². HPV11 also possesses a GRE consensus sequence and it is



The intact E6-E7 open reading frames which are transcribed in cells, such as HeLa cells, derived from cervical cancer tissue. The middle and bottom transcripts contain a spliced E6. The donor and acceptor signals for splicing are only found in HPV16 and 18 and not in HPV6 or 11. (Data taken from ref. 14.)

assumed that this region can also be activated by dexamethasone, although Pater *et al.* did not carry out point-mutation studies on either HPV11 or HPV16 DNA to ensure that these regions are indeed necessary for the transformation activity. In addition, it is important to note that the physiological and pharmacological activities of glucocorticosteroids are very different from oestrogens and progesterones, even though there is evidence that they may act via similar DNA consensus sequences.

HPV16 and 18 can transform a range of primary rodent cells including rat embryo fibroblasts and baby rat kidney cells^{3,4} in cooperation with activated *ras* oncogene, but not with activated *c-myc*. In some experiments, strong exogenous promoters, such as mouse moloney leukaemia virus long terminal repeats, are necessary to achieve transformation. Pater *et al.* look at HPV11 and 16 only under the control of the viruses' own endogenous promoter elements, and find that the presence of dexamethasone is necessary for transformed foci production in the kidney cells. Because different groups work with different plasmid constructions and target cells, comparison of results is not always easy.

In spite of these problems, it appears that the E6-E7 region (see figure) is implicated in transformation, and recent

work⁵ suggests that E7 is the major transforming gene. Evidence for this conclusion comes from the fact that although the integration pattern of the HPV genome into cellular DNA is random, the E6-E7 region is consistently transcribed in cervical tumour cell lines. Three transcripts from this region have been detected in cells grown from fresh cervical tumour tissue such as HeLa (containing 10 genome equivalents of HPV18 DNA) and Caski (containing 600 genome equivalents of HPV16 DNA) cells (see figure). E6* is predicted to be transcribed only in high oncogenic types such as HPV16 and 18 but not from low oncogenic types such as HPV6 and 11, as these latter viruses lack donor/acceptor sequences in the middle of E6. E6* may allow a more efficient translation of E7, as with the colinear transcript of E6-E7 no introns have been detected. Thus, the E6* transcript may be an important difference between HPV types, although a note of caution is warranted as an E6* product of 6.5 kilobases has been detected in mouse xenografts of HPV18-containing cervical tumour cells⁶, making it unlikely that there is an E7 product from this transcript. It is therefore still not clear which transcript will encode an E7 gene product.

The E6-E7 region is transcribed in the normal replicative cycle of HPV, so what is altered in the malignant cell? First, the genome is integrated and second, to ensure transcription of the E6-E7 region, integration usually occurs in the E1 or E2 open reading frames, uncoupling the poly(A) signals of the viral mRNA from the 5' end and resulting in use of cryptic signals in host DNA⁷. Further disruption in viral control mechanisms occurs because the E2 open reading frame encodes at least two proteins: one transcribed from a full-length transcript can *trans-activate* and upregulate transcription of other open reading frames⁸; and a shorter one transcribed from the carboxy half of E2 competitively inhibits the large product⁹. Integration may therefore result in uncontrolled transcription of E7 and affect the balance of virus-cell interaction, leaving these cells susceptible to other events that could lead to a malignant phenotype.

How does the E7 gene product function? Recently, this gene has been shown⁵ to

1. Pater, M.M. *et al.* *Nature* **335**, 832-835 (1988).
2. Gloss, D. *et al.* *EMBO J.* **6**, 3735-3743 (1987).
3. Storey, A. *et al.* *EMBO J.* **7**, 1815-1820 (1988).
4. Chesters, P.M. & McCance, D.J. *gen. Virol.* (in the press).
5. Phelps, W.C. *et al.* *Cell* **53**, 539-547 (1988).
6. Schneider-Gadicke, A. *et al.* *Cancer Res.* **48**, 2969-2974 (1988).
7. Durst, M. *et al.* *J. gen. Virol.* **66**, 1515-1522 (1985).
8. Phelps, W.C. & Howley, P.M. *J. Virol.* **61**, 1630-1638 (1987).
9. Chin, M.T. *et al.* *J. Virol.* **62**, 2994-3002 (1988).
10. Whyte, P. *et al.* *Nature* **334**, 124-129 (1988).
11. McCance, D.J. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 7169-7173 (1988).
12. Brinton, L.A. *et al.* *Int. J. Cancer* **38**, 339-344 (1986).
13. Peters, R.K. *et al.* *J. natn. Cancer Inst.* **77**, 1063-1077 (1986).
14. Schwarz, E. *et al.* *Nature* **314**, 111-113 (1985).

trans-activate the E2 gene of adenovirus, a function normally of adenovirus E1a. Two domains shared by E1a transcripts 12S and 13S, which are important for its co-operation with *ras* in the transformation of primary rat cells, have similar amino-acid sequences to an area of the E7 product of HPV16. This could be significant in the light of the finding that E1a interacts with the retinoblastoma (RB) gene product and that the complex can be immunoprecipitated from adenovirus transformed cells with E1a antisera¹⁰. The RB protein, a tumour-suppressor gene, is thought to be involved in cell-signalling pathways, so it may be an important component of cell differentiation or cell-cycle regulation. Removal of this gene by complexing with a viral-coded product could disrupt normal control mechanisms, which would result in a reduced capacity

to differentiate and greater potential to replicate.

HPV16 can immortalize and inhibit the differentiation of human epithelial cells *in vitro*¹¹. The cells could then be susceptible to other factors resulting in a cell capable of invasion. Oral contraceptives, as mentioned above, have been implicated in cervical cancer. The epidemiological evidence is contradictory^{12,13} and the relative risk may be due to other confounding factors, a sentiment shared even by those whose studies show an increased relative risk. It is not therefore possible to justify damning the pill on the evidence available when pregnancy may hold a greater risk to a woman's health than the use of oral contraceptives. □

D. J. McCance is in the Department of Microbiology, UMDS, Guy's Hospital, London Bridge, London SE1 9RT, UK.

Cosmology

A deep look at forming galaxies

Joseph Silk and August Evrard

YOUNG galaxies provide a test for theories of galaxy formation, a process that was the subject of a recent meeting*. Although no consensus emerged on when galaxies formed, the confrontation of new discoveries of very distant and possibly young galaxies with theoretical speculations about forming galaxies led to a stimulating exchange of ideas.

Protogalaxies have been the holy grail of observational astronomy for two decades, but the meeting was especially timely as it marked the discovery during the past year of objects that are tantalizingly close to being confirmed as protogalaxies. The strongest claim was made by L.L. Cowie (Institute of Astronomy, Hawaii) and collaborators who have obtained optical and infrared CCD (charge-coupled device) images — down to 27 magnitude in the visible band, about as faint as previously obtained by A. Tyson (AT&T Bell Labs) in a similar deep survey of arcminute-square fields at high galactic latitude. The novel aspect of the new work is that images in several colours reveal among the many faint galaxies a few which have flat spectral energy distributions.

Cowie and colleagues infer from the energy distributions that these objects are forming galaxies undergoing an extreme burst of star formation at redshift $z = 2.8$ – 3.6 , corresponding to perhaps 8 – 16×10^9 years ago. The total flux from these objects corresponds to a substantial fraction of that emitted by the early generations of stars that synthesized the metals seen in galactic disks.

Dramatic evidence for galaxy formation at high redshift comes from the optical

identification of known luminous radio galaxies. A radio galaxy at $z = 3.395$ discovered by S.J. Lilly (Institute for Astronomy, Hawaii; see the recent News and Views article by R.F. Carswell, *Nature* **335**, 119; 1988) seems from its visible and near-infrared colours to be forming new stars and to contain a substantial population of stars that formed at redshift 5 or before. This conclusion arises because, for the simplest models of galactic evolution to be valid, the infrared emission detected at a wavelength of 2 micrometres with this redshift must have been formed 1 – 2×10^9 years before $z = 3.395$.

Elongated distribution

Lyman- α emission from this radio galaxy (H. Spinrad, University of California, Berkeley) has an elongated distribution extending over 100 kiloparsecs with a surface brightness resembling that of a giant disk. This object and at least seven other emission-line radio galaxies beyond $z = 2$ are elongated along the main axes of the associated radio-emitting regions. This contrasts with low-redshift galaxies whose radio lobes tend to lie along the minor axes of luminous elliptical galaxies (McCarthy, P. *et al.* *Astrophys. J.* **321**, L39–L44; 1987). The implications of the apparently causal relations between these emissions remain to be explored.

The low intergalactic neutral hydrogen (H I) column density inferred from quasar absorption lines at $z \leq 4$ — the 'Gunn–Peterson' constraint — also indicates star formation at high redshift. P. Shapiro (University of Texas) presented a model in which star formation starts at $z \geq 5$ that provides sufficient ionizing flux to satisfy

the Gunn–Peterson limit. One consequence is that half the observed metal abundance of population I (young, metal-rich) stars must have been produced before $z = 3$. This must be reconciled with the interpretation of the objects at $z \approx 3$ discovered by Cowie and co-workers that can account for almost all of the early star formation in massive galaxies (see above).

High-redshift systems

A. Wolfe (University of Pittsburgh) and collaborators have indirectly discovered protodisks at $z = 2$ – 3 by studying absorption spectra imposed on the emission of distant quasars. They interpret the absorption lines as broad (or 'damped') redshifted Lyman- α profiles appropriate to intervening disks of neutral hydrogen with column densities greater than 10^{20} cm⁻². The high-redshift systems occur so frequently that they cover about a fifth of the sky, and contain perhaps as much gas as contained in stars within present galactic disks. In one case illuminated by a quasar with extended radio emission, the neutral-hydrogen cloud has a radius of about 100 kiloparsecs, the putative scale of a protodisk. And the profiles of the 21-centimetre absorption lines in several cases have widths comparable to those seen in thin galactic disks.

Weak, narrow Lyman- α emission superposed on a system of damped Lyman- α emission at $z = 2.5$ has been observed (R. Hunstead and M. Pettini, Anglo-Australian Observatory). It seems that the rate of star formation is surprisingly low if this is a galactic disk that is predominantly gaseous and is yet to form most of its stars. If there are protodisks at $z = 2$ – 3 , we would expect also to find protospheroids forming at higher redshift as spheroids have older stellar populations than disks.

A systematic search for optical counterparts to steep-spectrum (thus remote) faint radio galaxies (K.C. Chambers, Johns Hopkins University; G.K. Miley, Space Telescope Science Institute; W.J.M. Breughel, University of California, Berkeley) has revealed six at $z > 2$, it was reported at a separate meeting[†]. These include the current distance record holder at $z = 3.8$. The systems are dominated by line emission from hot gas and the inferred rate of formation of massive stars required to excite the gas is comparable to that expected for a luminous protogalaxy. Alternatively, the gas may be ionized by radiation from the nucleus, although only improved observations will discriminate between these possibilities.

Assessing the relevance of radio galaxies to galaxy formation is made hard by our lack of knowledge of how significant a proportion they are of all high-redshift galaxies. For example, cold-dark-matter

* The Epoch of Galaxy Formation, Durham, 18–22 July 1988.

† Annual assembly of the International Astronomy Union, Baltimore, 2–11 August 1988.

theory, discussed by many participants, predicts that most stars formed at low redshift but permits some galaxies to form back to $z = 5$ and bare galactic nuclei even earlier. These predictions are sensitive to the degree of 'biasing' (the relative clustering strength of visible and dark matter) incorporated. Rival theories involving cosmic strings (R. Brandenburger, Brown University) or explosions triggered by the decay of superconducting cosmic strings (C. Thompson, Princeton University), are still too 'soft' to make testable predictions for observational astronomers.

The best prospect for resolving the

uncertainties is the analysis of the deep optical redshift surveys being performed on square-degree fields by D. Koo (University of California, Santa Cruz) and R. Kron (Chicago University) and by R. Ellis and collaborators (Durham University). Preliminary results for these deep, optically unbiased, large-scale surveys, which include redshifts of several hundred objects down to 21 magnitude, show that extensive evolution is already occurring at a redshift of about unity. □

Joseph Silk and August Evrard are in the Department of Astronomy, University of California at Berkeley, Berkeley, California 947200, USA.

Biomechanics

Muscle function in locomotion

Ian Johnston and John Altringham

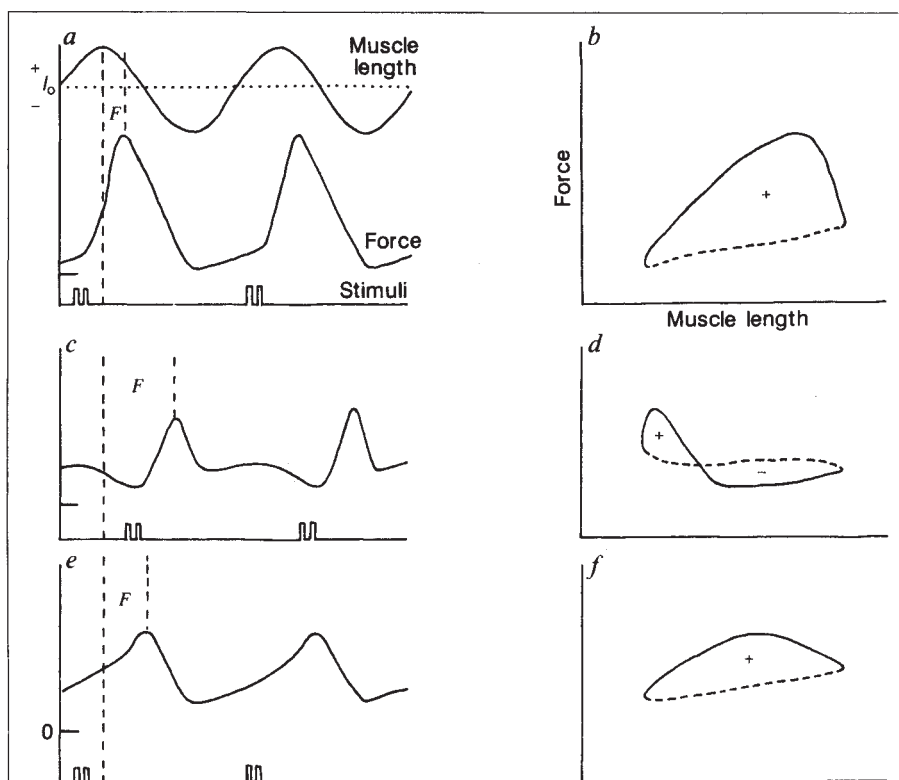
How is animal locomotion related to the properties of individual elements of the musculo-skeletal system? Few studies have successfully bridged the gap between muscles and locomotion, although it has long been recognized that animals possess a range of muscle fibre types specialized for particular functions¹. In one attempt to attack this problem, Lawrence Rome and co-workers on page 824 of this issue² ask whether the different fibre types in a swimming fish are operating under conditions of maximal power output and efficiency.

Fish muscle is an ideal preparation to use in addressing this problem because in most other animals it is not known which fibres are active during particular kinds of movement. Rome *et al.* take advantage of the fact that in fish there is distinct anatomical separation and division of labour of muscle fibre types within each body segment or myotome. Slow fibres in the carp form a superficial ribbon, directly beneath the skin, running almost parallel to the spinal cord. The orientation of fast fibres, projected through successive myotomes, follows helical trajectories⁴. Rome *et al.* calculate sarcomere length excursions from changes in body flexure digitized from high-speed cine film. At low speeds, they find that slow fibres shorten at velocities yielding maximum power output and peak efficiency, which were calculated from the force-velocity relation of isolated muscle fibres (0.7–1.7 muscle lengths per second, 1 s^{-1}). But during fast starts, Rome *et al.* suggest that the slow fibres would need to shorten at four times their maximum contraction velocity ($V_{\max} = 4.71 \text{ s}^{-1}$), and greater than the $V_{\max}$ of the fast fibres (13.1 s^{-1}). The authors calculate that because of the complex myotomal geometry, fast fibres need shorten at around only 40 per cent of their maximum shortening velocity, which they postulate to be within the range for maximum

power output.

Further insight into the functional significance of the complex fibre geometry of fish myotomes is provided by kinematic studies on the saithe. Hess and Videler show⁵ that during slow, continuous swim-

ming, the length changes of superficial muscle in this species are essentially sinusoidal, involving sarcomere length excursions from around 1 per cent behind the head to around 6–7 per cent near the tail. Their analysis indicates that the bending moment is a standing wave, and they predict essentially simultaneous muscle activity on alternate sides of the body. Because the wave of body deformation runs from head to tail, this results in systematic phase differences between force and shortening/lengthening velocity. At a distance of 0.65 body lengths from the head, force is maximal as the fibres shorten through their resting length, such that force generation and power output are positive through most of the tail-beat cycle. Power output decreases in both caudal and rostral directions, as the phase relation changes. In the caudal region, maximum force is generated just after the fibres are maximally stretched. Thus, if all fibres of a given type are to operate optimally, some regional variation in their properties would be predicted. For fast fibres, this is likely to be true between anterior and posterior myotomes, and between superficial and deep



Idealized representation of oscillatory work performance by muscle. *a, c, e*, Muscle length, force and stimuli plotted against time for two different phase angles, F = per cent cycle time between peak tension and maximum length. *b, d, f*, Force versus length loops, where the loop area is net work per cycle. *a, b*, $F = 12.5$ per cent, peak force generation occurs during shortening from maximum through rest length, l_0 (dotted line), and the muscle is restretched when force is low. A large loop of positive work is performed over the whole cycle. *c, d*, $F = 40$ per cent, peak force occurs near minimum length, and the muscle is restretched when tension is still high. A work loop is generated with both positive and negative components. *e, f*, Stimulation at same time as in *a*, but muscle has slower rates of tension rise and relaxation than in *a*. Net work is reduced by the lower peak force, and the need to restretch the muscle when still active. (Adapted from ref. 6.)

layers, because some superficial fibres have similar orientations to slow fibres.

Muscle power output is most commonly estimated from force-velocity curves determined from shortening under constant load, a situation rarely encountered *in vivo*. However, such measurements fail to take into account the work done on restretching muscle after shortening, the effects of previous contractions on the dynamic properties of the force-generating molecules, the crossbridges, or the fraction of time the muscle is active during each cycle. In an elegant study on synchronous insect flight muscle, Josephson measured⁶ muscle power output in a way that mimics the normal operating conditions *in vivo* (see figure). Josephson applied single or multiple stimuli to muscle fibres at selected phases of imposed sinusoidal length changes, at a frequency and amplitude appropriate to flight. By plotting force against length, 'work' loops can be generated, allowing the power output of the muscle to be determined by calculating work per cycle multiplied by frequency.

This approach has wide applicability, and Stevens has now used it⁷ to measure power output in frog sartorius muscle. When a frog hops, the main leg muscles are alternately stretched then shortened at around 2 cycles per second. The work done per cycle is highly dependent both on the number of stimuli and on their phase relation with the length-change cycle^{6,7}. Work increases with excursion amplitude, as a complex function of the force-velocity, length-force and isometric time-force relations, to a maximum at ± 6 per cent of the fibre's rest length⁷. This corresponds to the actual range of excursion amplitudes found *in vivo* from measuring the changing angle of each joint as the leg extends in jumping⁸. For muscles operating under these conditions, it is clear that contraction kinetics and cycle frequency must be 'matched' to produce maximum power output (see figure). In view of this, an important experiment would be to determine the way in which power output of fast and slow muscles varies with cycle frequency. In fish, for example, one might expect the frequency for maximum efficiency of slow muscle fibres to coincide with the most economical cruising speed. As swimming

Jean Brachet (1909–1988)

ORIGINAL scientific ideas, once verified and accepted, become commonplace and once-pioneering lines of investigation become established fields. One of these pioneers, Jean Brachet, died on 10 August. It was the provocative data and suggestions of Brachet which initiated the intense research activity that centred on RNA during the



Jean Brachet — a scientific pioneer.

1950s, and contributed much to the emergence of the idea that DNA led to RNA led to proteins.

In 1929, Brachet, a 20-year-old medical student, set out to identify the molecules involved in morphogenesis. A simple but well-mastered histological method allowed him to demonstrate that thymonucleic acid (now known as DNA) is a constant constituent of cells. Using the Feulgen reaction and the diphenylamine method, Brachet demonstrated that sea urchin eggs are the site of considerable DNA synthesis during cleavage and development. Soon after, by combining staining by basic dyes and treatment of tissue sections by crystalline ribonuclease, he showed that basophilia of cytoplasm is due to RNA. As a consequence, RNA, previously considered as an exotic compound of yeast, was shown to exist in the nucleolus and the cytoplasm of all cells. The following year, polytene chromosomes of *Drosophila* and *Chironomus* were shown to contain different amounts of RNA.

speed increases, the sequential recruitment of fast oxidative and fast glycolytic fibres⁹ (with shorter twitch contraction times) is needed to match the increased tail beat frequency.

The maximum force per unit cross-sectional area and maximum relative shortening are governed by the dimensions of the muscle's contractile filaments and are independent of body size. The amount of work done per muscle-fibre contraction is therefore a scale-independent variable, and power output is proportional to contraction frequency¹⁰. Maximum tail beat frequencies range from 40–60 Hz in 1-cm fish to 5–8 Hz in 2-m fish. Because the distance moved in one tail beat is a nearly constant proportion

The systematic and quantitative comparison of the basophilia (removed by ribonuclease) with the pentose content of the cells, led Brachet to the following observations: that cells actively engaged in protein synthesis have high amounts of RNA; that RNA is found in the nucleolus, and, during embryonic development, increases before cytoplasmic RNA; and that RNA is found in the 'ergastoplasm' (the endoplasmic reticulum, where active protein synthesis was suspected to occur). Brachet concluded that RNA is most probably involved in protein synthesis, while independently, Caspersen in Sweden reached similar conclusions.

Just a few years elapsed before RNA-rich microsomes, discovered by Albert Claude in tissues of Rous sarcoma virus-infected chicken, were identified by Brachet and Jeener in all mammalian cells examined. Their fundamental finding was that haemoglobin is specifically associated with red blood cell granules, amylase with salivary gland particles and trypsin with pancreatic particles. Their conclusion in 1944 that "to the ribonucleoproteins could be assigned the role of essential agents of protein synthesis" was brilliantly confirmed by Zamcenik's group when, in 1958, they proved that microsomes were the sites of incorporation of ¹⁴C amino acids into proteins.

Brachet also wrote a series of influential books which informed a wide audience about the rise of the new field: *Chemical Embryology* in 1950, *Biochemical Cytology* in 1957, and *Molecular Embryology* in 1974. It is a measure of Brachet's drive that he remained to the end as involved and stimulating as he had always been.

As well as his scientific contribution, Brachet will be remembered by his many students and friends as a true humanist in his deep respect for life, his sincere approach to ethical issues, and his incessant fight for freedom.

A. Burny

A. Burny is a professor in the Department of Molecular Biology, University of Brussels, 1640 Rhode-Saint-Genève, Belgium.

of body length¹¹, tail beat frequency is a major factor limiting maximum length-specific swimming speed. Tail beat frequency is strongly correlated with muscle contraction kinetics (activation and relaxation times)¹². However V_{max} , which is just one determinant of contraction kinetics, is largely independent of size¹³. The slow fibres of small animals can have shorter twitches than the fast fibres of large ones. The problems associated with scaling¹⁴ provide a challenging test for this approach to modelling muscle power output during locomotion. □

Ian Johnston and John Altringham are in the Gatty Marine Laboratory, Department of Biology and Preclinical Medicine, St Andrews University, St Andrews, Fife KY16 8AQ, UK.

1. Humphrey, G.M. *J. anat. Physiol.* **6**, 293–376 (1872).
2. Rome, L.C. *et al. Nature* **335**, 824–827 (1988).
3. Bone, Q. in *Fish Physiology*, Vol. VII, Ch. 6 (eds Hoar, W.S. & Randall, D.J.) (Academic, New York, 1978).
4. Alexander, R. McN. *J. mar. Biol. Assoc. U.K.* **49**, 263–290 (1969).
5. Hess, F. & Videler, J.J. *J. exp. Biol.* **109**, 229–251 (1984).
6. Josephson, R.K. *J. exp. Biol.* **114**, 493–512 (1985).
7. Stevens, E.D. *J. Muscle Res. Cell Motil.* **9**, 329–333 (1988).
8. Calow, L.J. & Alexander, R. McN. *J. Zool.* **171**, 293–321 (1973).
9. Johnston, I.A. *et al. J. comp. Physiol.* **114**, 203–216 (1977).
10. Schmidt-Nielsen, K. *Scaling: Why is Animal Size so Important?* (Cambridge University Press, 1984).
11. Webb, P.W. *et al. J. exp. Biol.* **109**, 77–95 (1984).
12. Wardle, C.S. *Nature* **255**, 725–727 (1975).
13. McMahon, T.A. *J. appl. Physiol.* **39**, 619–627 (1975).
14. Hill, A.V. *Sci. Prog.* **38**, 209–230 (1950).

Metamorphic processes

Mantle melts in diamonds

Martin A. Menzies

DIAMOND-BEARING kimberlites and other alkaline magmas erupted at the surface are believed to have formed at great depths in the Earth's mantle, beneath the lithospheric plate that carries the continents. Their journey to the surface can involve entrainment of high-pressure minerals, like diamond, and peridotite xenoliths from the lithosphere¹. These magmas, rich in volatile and trace elements, can exchange material with (or metasomatize) the mantle wall-rock that surrounds the conduit as the magmatic plumbing is established through the lithospheric shell. Undeniable petrographic and chemical evidence² has accumulated for such an interaction in the form of metasomatism via silicate melts (in which dissolved oxides dominate volatiles) or fluids (in which volatiles dominate dissolved oxides). The discovery of highly potassic, volatile-rich micro-inclusions in diamonds (see Fig. 1) from Zaire and Botswana, reported by Navon *et al.* on page 784 of this issue³, has important implications for our understanding of such mantle processes.

These assemblages of hydrated sheet silicates (such as mica), carbonate (calcite) and apatite, with enhanced water and carbon dioxide content, point to the existence of trace-element- and volatile-enriched oxidizing fluids or melts 150 km below the surface, near the base of the lithosphere. Because diamond is stable for certain pressures and temperatures only, corresponding to a depth of 150–180 km, any solution to the origin of these diamond micro-inclusions must appeal to processes that were active at or below that level—in the conductively cooling lithosphere or the convecting asthenosphere. Four possible origins arise for the micro-inclusions (see Fig. 2).

First, they could be derivatives of a potassic magma such as lamproite, as suggested by Navon *et al.*³. The highly potassic nature of lamproites and certain exotic mantle xenoliths encouraged Navon *et al.* to suggest a genetic link between the formation of diamonds (plus micro-inclusions) and the formation of 'marid' (mica-apatite-rutile-ilmenite-diopside) pegmatites (igneous veins) in the lithosphere. Marid xenoliths are believed to represent derivatives of lamproites⁴ or kimberlites that failed to erupt. Because magma heat death is thought to be an important mechanism responsible for mantle metasomatism⁵, it is not surprising that the crystallization of

marid pegmatites in the lithospheric mantle is associated with the production of evolved fugitive melts that crystallized as mica and amphibole in the surrounding mantle. Marid xenoliths, however, do not bear diamond (unlike eclogites and peridotites) mainly because they formed within the amphibole stability field at pressures considerably lower⁵ than the diamond stability field. Clearly, the clue

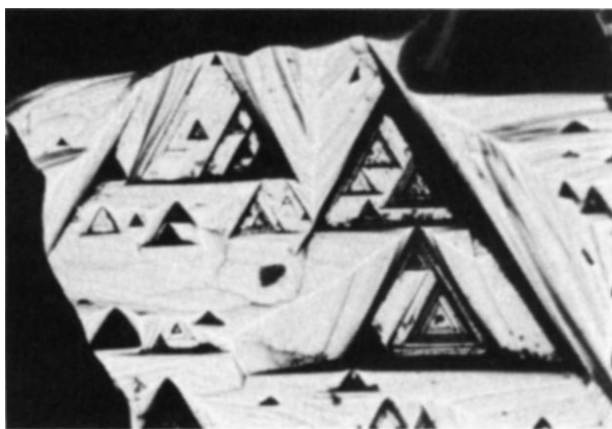


Fig. 1 Etched pits on a coated diamond recovered by Navon *et al.*

to understanding the origin of these micro-inclusions is to be found at higher pressures.

Second, the micro-inclusions could represent a melt enriched in light rare-earth elements, similar to that involved in the production of Archaean calcium-poor garnets found as inclusions in megacrystic diamonds. Calcium-poor garnets enclosed in diamonds⁶ from South Africa provided the first unequivocal evidence that the sub-cratonic or sub-Archaean lithospheric mantle was old and thick and that in the Archaean it projected into the diamond stability field. Moreover, the rare-earth geochemistry of these garnet inclusions indicates that Archaean enrichment processes (such as metasomatism) had been active at or below the depth of the diamond stability field. Although the formation of the peridotitic inclusions in diamonds and the micro-inclusions both involved melts or fluids enriched in the light rare earths, a close temporal and spatial relationship is rather unlikely. This is because the garnet inclusions have strontium and neodymium isotope ratios somewhat different from the micro-inclusions⁷ possibly because they have different ages. Until the age of the micro-inclusions is known, the evidence for a genetic link will be, at best, equivocal.

Third, the micro-inclusions could be a metasomatic fluid fortuitously captured during the crystallization of the diamonds.

Mantle metasomatism, which is roughly a geological ion-exchange process, is a convenient mechanism whereby the trace-element and volatile inventory of an otherwise refractory rock can be substantially increased by the growth of new minerals. Consequently, in the early 1980s it was uncritically embraced as a panacea for any problems concerning the origin of basaltic magmas. Metasomatism has been reported from Archaean⁸ Proterozoic⁹ and Phanerozoic¹⁰ lithospheres and as such seems to have been active within the Earth's mantle for more than three thousand million years. Because metasomatism is inextricably linked to the thermodynamically irreversible ascent of

volatile-bearing magma through cracks in the lithosphere, metasomatic minerals (mica, apatite and carbonate), like those found within the micro-inclusions³, are invariably rich in incompatible elements like the light rare earths, rubidium and potassium.

The metasomatic fluid can also be rich in water or carbon dioxide depending on the location of the carbonate stability field and the vapour-present curve¹¹. Recent experimental work^{12,13} and geothermobarometric studies¹⁴ of garnet peridotites indicate that metasomatism occurs at many pressures as well as in the diamond stability field. The micro-inclusion data³ confirm the existence of melts and fluids with up to 50 per cent volatiles at depths of 150–180 km. Such potassic volatile-rich materials clearly

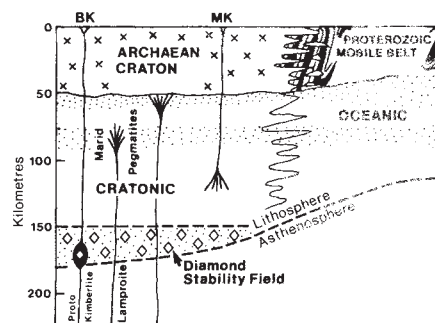


Fig. 2 Hypothetical cross-section of Archaean and Proterozoic lithosphere¹. Archaean cratons (thick crust and low heat flow) are underlain by a heterogeneous sub-crustal lithospheric keel ('cratonic') that differs markedly from the homogeneous 'oceanic' lithosphere¹⁴ beneath adjacent Proterozoic mobile belts (thin crust and high heat flow). Sub-Archaean lithosphere is so thick⁶ that it projects into the diamond stability field. Consequently most diamond-bearing kimberlites are erupted in cratonic areas and not circumcratonic mobile belts. The diamond stability field is traversed by several asthenospheric melts *en route* to the surface. As a result, basaltic kimberlite (BK) can transport diamonds, be a source for 'phenocrystal' coating on diamonds and be a powerful metasomatic agent should it fail to erupt. (Stippled pattern, metasomatic levels; MK, micaceous kimberlites).

could be powerful metasomatic agents.

The last possibility is that the micro-inclusions could be derived from proto-kimberlite — the postulated host magma that brought the diamonds to the surface. Because the upward migration of asthenospheric magmas through the lithosphere is believed to be responsible for many forms of metasomatism, perhaps these micro-inclusions originated at greater depth, in the underlying asthenosphere. Basaltic kimberlites, such as those which entrained the Botswana diamonds, came from the asthenosphere (or deeper) as did ocean island basalts¹⁰. This contrasts markedly with micaceous kimberlites which seem to have formed primarily in the lithosphere. In the case of the Zaire diamonds, other investigations point to a close genetic relationship between the micro-inclusion bearing diamonds and kimberlite. Chondrite-normalized rare-earth abundances and strontium-isotope data⁷ for the micro-inclusions within Zaire diamonds fall in the range of basaltic kimberlites. Furthermore, the coatings on several Zaire diamonds have uniform carbon and nitrogen isotope compositions¹¹ which point to an origin in a homogeneous rather than a heterogeneous mantle reservoir.

The sub-Archaean lithosphere is extremely heterogeneous for radiogenic and stable isotopes whereas the underlying asthenosphere is relatively homogeneous and in many ways similar to the sub-Proterozoic lithosphere. If the host kimberlitic or proto-kimberlitic magma is asthenospheric in origin the carbon and nitrogen data could indicate that the coated nature of diamonds results from 'phenocrystal overgrowth' on a xenocrystic core¹². Therefore the micro-inclusions could represent phases that crystallized from melt or fluid derivatives of proto-kimberlite incorporated in the diamond coat.

As such the metasomatic-fluid hypothesis outlined above is inextricably linked to any hypothesis involving proto-kimberlite. One of the main arguments³, however, against the micro-inclusions being a derivative of kimberlite is the

higher concentration of potassium and volatiles in the potassic inclusions, a factor that does not favour direct comparison with the composition of erupted kimberlites. The erupted product may, however, have changed significantly since it was 150–180 km deep. In a recent text on kimberlites¹³, Mitchell concluded that "a unique drawback to devising a petrogenetic model [for kimberlites is] our complete ignorance of the composition of primitive kimberlite magma".

The discovery of water- and carbon-dioxide-rich micro-inclusions in diamonds from 150–180 km contributes significantly to elucidating the distribution of volatiles and the nature of fluids and melts at high

pressures. Moreover, their occurrence may help resolve some of the highly contentious issues concerning the oxidation state of the lithosphere. The micro-inclusions³ could be the first recorded sample of proto-kimberlite, or a derivative thereof, entombed in a diamond time capsule. The enhanced trace-element and volatile composition can indicate either that our understanding of elemental partitioning at depths of 150–180 km is in its infancy or that the chemistry of proto-kimberlite is very different from the kimberlite erupted at the surface. □

Martin A. Menzies is in the Department of Geology, Royal Holloway and Bedford New College, Egham, Surrey TW20 0EX, UK.

Cellular neurobiology

Synaptic memory molecules

Mary B. Kennedy

the thing I came for:
the wreck and not the story of the wreck
the thing itself and not the myth¹

MOST neuroscientists believe that memories are stored via highly selective changes in the strength of synaptic connections between neurons in the brain. Increases in synaptic strength, for example, may reinforce the pattern of neuronal activity during a remembered event so it can be more easily reactivated later. Understanding synaptic plasticity is the first step in making accurate network models of memory formation. One form of plasticity, long-term potentiation (LTP), a long-lasting increase in the effectiveness of synaptic transmission^{2–4}, is triggered by brief, high-frequency stimulation of certain presynaptic axons (Fig. 1). It is readily demonstrated in the hippocampus which plays a crucial role in the formation of memories⁵. Two new papers help to unravel the molecular basis of LTP.

One, by Kauer *et al.*⁶, demonstrates that LTP can be separated into two processes: a relatively short-term, slowly decaying potentiation (SDP) that disappears within about 30 minutes; and a true LTP that lasts for hours or days. They induce slowly decaying potentiation by applying glutamate, an excitatory neurotransmitter, over the surface of the synapses. Establishment of true LTP requires in addition electrical stimulation of the presynaptic terminals, implying that the release of a second, unknown, neurotransmitter is needed. In the other study, reported by Malinow *et al.* on page 820 of this issue⁷, the authors show that one or more constitutive protein-kinase activities are involved in maintaining true LTP, but seem not to be involved in the less persistent SDP. The events underlying LTP in the hippocampus are thus coming into sharper focus (Fig. 2).

LTP has traditionally been divided into two phases; the first, induction, includes several seconds during and after tetanic stimulation when the events that trigger LTP begin. The second, maintenance, begins when the final level of increased synaptic efficacy has been reached and is characterized by a sustained alteration of the presynaptic terminal that causes it to release more excitatory neurotransmitter during a single presynaptic stimulus, producing a larger postsynaptic depolarization⁸. There may also be a sustained postsynaptic change enhancing the 'coupling' between depolarization of postsynaptic cells and generation of action potentials⁹.

At most hippocampal synapses, the induction phase seems to involve a curious interplay between pre- and postsynaptic mechanisms. Forced hyperpolarization of the postsynaptic cell by voltage clamp during tetanic stimulation blocks the induction of LTP¹⁰. Thus, although the maintenance phase is primarily expressed in the presynaptic terminal by increased release of neurotransmitter, part of the induction process clearly occurs postsynaptically. This presents an interesting mechanistic puzzle as there are no other examples of highly localized feedback from a postsynaptic site onto its own presynaptic terminal. The complex interaction underlying induction is likely to involve several molecular steps, and interest has centred on identification of the steps that initiate the synaptic change and the molecules that remain altered during maintenance, thereby serving as sites of information storage.

A significant result was the demonstration that induction of LTP requires activation of a subset of excitatory glutamate receptors called NMDA receptors¹¹. There are two pharmacologically distinct types of glutamate receptors in the brain,

1. Nixon, P. ed. *Mantle Xenoliths* (John Wiley and Sons, Chichester 1987).
2. Menzies, M.A. & Hawkesworth, C. J. *Mantle Metasomatism* (Academic, London, 1987).
3. Navon, O., Hutcheon, I.D., Rossman, G.R. & Wasserburg, G.J. *Nature* **335**, 784–789 (1988).
4. Waters, F.G. *Contr. Mineral. Petrol.* **95**, 523–533 (1987).
5. Waters, F.G. & Erlank, A.J. in *Oceanic and Continental Lithosphere – similarities and differences* (eds Menzies, M.A. & Cox, K.G.) 185–204 (Oxford University Press, 1988).
6. Richardson, S.H., Gurney, J.J., Erlank, A.J. & Harris, J.W. *Nature* **310**, 198–202 (1984).
7. Akagi, T. & Masuda, A. *Chem. Geol.* **70** (1), 2 (1988).
8. Green, D.H. & Wallace, M.E. *Nature* (in the press).
9. Wyllie, P.J. in *Mantle Xenoliths* (ed. Nixon, P.) 609–622 (John Wiley and Sons, Chichester, 1987).
10. Smith, C.B. *Nature* **304**, 51–54 (1983).
11. Boyd, S.R. & Pillinger, C.T. *Chem. Geol.* **70** (1) 46 (1988).
12. Boyd, S.R. *et al.* *Earthplanet. Sci. Lett.* **86**, 341–355 (1987).
13. Mitchell, R.H. *Kimberlites* (Plenum Press, New York, 1986).
14. Boyd, F.R. & Mertzman, S.A. in *Magmatic Processes and Physicochemical principles* (ed. Mysen, B.) 13–24 (American Geophysical Union, Washington DC, 1987).

quisqualate/kainate (Q/K) and *N*-methyl-D-aspartate (NMDA). Blockade of the former eliminates the postsynaptic potential normally produced by a single synaptic stimulation. In contrast, specific blockade of NMDA receptors does not diminish 'normal' synaptic transmission, but can completely block induction of LTP during tetanic stimulation. Strong depolarization of the membrane and the presence of glutamate are both required to activate NMDA receptors¹², therefore tetanic stimulation is necessary for induction of LTP in a single pathway. This property of the receptor could provide a cellular mechanism for associative learning¹³.

How does activation of NMDA receptors trigger processes that contribute to LTP? The ion channel linked to Q/K receptors is highly selective for sodium and potassium ions, whereas the channel linked to NMDA receptors also admits calcium ions¹⁴. Activation of NMDA receptors can thus produce a local increase in the intraneuronal concentration of calcium, known to be second messenger that can activate different effector molecules. Three general mechanisms for LTP invoking calcium-activated enzymes have been proposed, all supported by some experimental evidence, but all with flaws. At present, these hypotheses can be tested only in limited ways.

Two of the hypotheses invoke protein kinases, which catalyse the transfer of phosphate to the side chains of certain amino-acid residues within proteins. Phosphorylation often alters the folding of proteins and thereby their function. Many ion channels and receptors in neurons are regulated reversibly via protein phosphorylation¹⁵.

One of the proposed mechanisms for LTP suggests that it is triggered and/or maintained by activation of a C-kinase via calcium ions and/or the second messenger diacylglycerol. C-kinases can be activated artificially by certain phorbol esters that dissolve in the cell membrane and act as diacylglycerol analogues¹⁶. The first indication that the C-kinase might be involved in LTP was the discovery that application of phorbol ester to a hippocampal slice produces a dramatic potentiation of synaptic transmission with many of the characteristics of LTP¹⁷. Akers *et al.* then reported¹⁸ that membranes of hippocampal neurons prepared from brains an hour after experimental induction of LTP contain twice as much C-kinase as membranes from unpotentiated neurons. They proposed that induction of LTP is accompanied by persistent activation of C-kinase as this enzyme is known to bind to membranes when activated. These experiments led to the idea that tetanic stimulation of hippocampal synapses activates C-kinase, which then phosphorylates specific proteins, producing LTP. Once such protein could be GAP-43

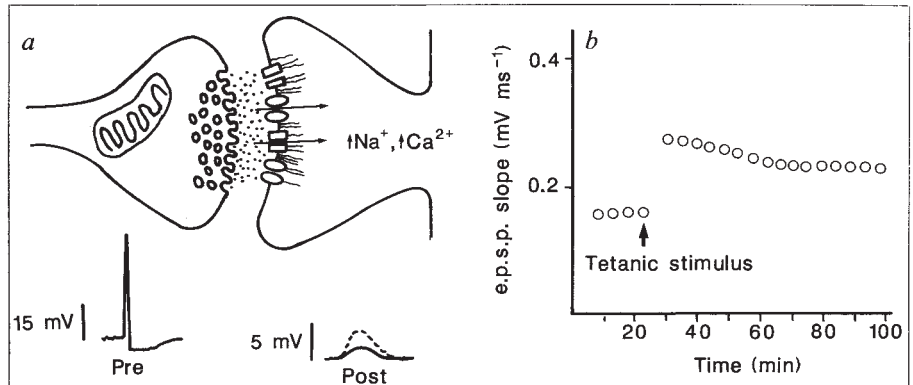


Fig. 1 Neurons in the brain receive thousands of synaptic contacts from other neurons. Each synapse consists of a 'presynaptic terminal' from the sending neuron and a specialized postsynaptic site, often in the shape of a tiny 1- μ m spine that protrudes from one of the many branch-like dendrites radiating from the neuronal cell body (*a*). When a presynaptic terminal is activated by an action potential (a large transient depolarization of the electrical potential across its membrane), small vesicles fuse with the presynaptic membrane and release a chemical neurotransmitter onto the adjacent postsynaptic membrane. The transmitter activates receptors that open ion-conducting channels. The resulting ion flow causes a small, brief (milliseconds) depolarization of the postsynaptic membrane, an excitatory postsynaptic potential (e.p.s.p.). The e.p.s.p.s produced by different synapses add together across the neuronal membrane. When many synapses are activated together, the total depolarization reaches a critical threshold, and an action potential is triggered that travels the entire length of the axon to activate the next synapses in line. LTP^{3,4} refers to a long-lasting increase in the size of the e.p.s.p. (dashed line). LTP is induced by a few brief, high-frequency (such as 100 Hz, 2s) bursts of presynaptic action potentials (tetanic stimulation). The increase in size of the e.p.s.p. is often measured as an increase in the rising slope of the extracellularly recorded e.p.s.p.s from several neurons (the population e.p.s.p.) (*b*). Larger e.p.s.p.s enhance the ability of the synapse to trigger action potentials in the postsynaptic neuron; this could be a memory-forming mechanism.

(also called F1 and B50; ref. 19), a prominent substrate for C-kinase in the brain.

Still missing from this hypothesis is an explanation of how phosphorylation of GAP-43 or any other protein phosphorylated by the C-kinase enhances release of neurotransmitter. In addition, because of the small size of synapses relative to cells, it is not yet technically feasible to investigate which proteins are phosphorylated in synapses during induction of LTP. The paper by Malinow *et al.* in this issue⁷ adds another fly to the ointment by reporting that potentiation of transmission by phorbol ester disappears within one hour after the drug is washed out of the synapses. Thus, activation of protein kinase C by phorbol ester is not sufficient to establish true LTP.

A second proposed mechanism for LTP involves the type II Ca^{2+} /calmodulin-dependent protein kinase (CaM kinase), which is activated when an increase in calcium concentration causes the Ca^{2+} -bound form of calmodulin to bind to a regulatory domain within the kinase. Indirect evidence suggests that this kinase is a target for the local calcium signal generated by activation of NMDA receptors. In the forebrain and hippocampus, where it is expressed at unusually high concentrations, the CaM kinase seems to be a major constituent of postsynaptic densities, cytoskeletal elements lying under postsynaptic membranes²⁰ and presumably exposed to NMDA receptor-mediated calcium flux. On activation *in vitro*, the kinase begins to phosphorylate appropriate substrate proteins, but it also

rapidly autophosphorylates itself on a threonine residue that lies next to the bound calmodulin²¹. When the calcium concentration subsequently falls, allowing the calmodulin to diffuse away, the autophosphorylated threonine prevents the refolding of the kinase that would otherwise lead to inactivation. This unusual property could, in theory, allow the CaM kinase to remain activated in neurons long after an initial calcium signal has decayed. The physical constraints on the accuracy and longevity of information storage by this switch-like molecule were explored recently²² in a mathematical model. Although the details of this model are almost certainly not accurate *in vivo*, it does provide a complete description of a potential mechanism for molecular information storage at single synapses. The CaM kinase hypothesis shares the weaknesses of the C-kinase hypothesis. In addition, there are no lipid-soluble activating agents with which to study physiological effects of CaM kinase activation.

The third hypothesis invokes irreversible proteolysis of synaptic molecules by calpains, calcium-activated proteases²³ in synaptic membranes which cleave the cytoskeletal proteins fodrin and MAP2. When applied to intact neurons, NMDA stimulates calpain activity²⁴. *In vitro*, calpain can proteolyse several protein kinases, including both C-kinase²⁵ and CaM-kinase (A. Kwiatkowski and M. King, unpublished observations), freeing the catalytic domain and producing constitutively active kinase molecules that could persist *in vivo* until destroyed by

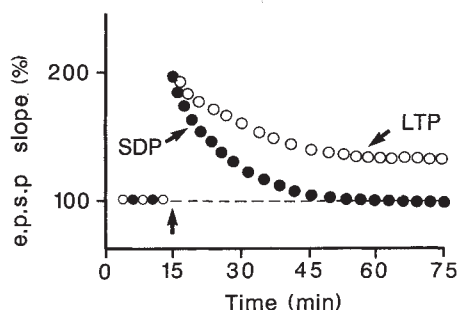


Fig. 2 Kauer *et al.*⁶ show that hippocampal synapses display two mechanistically distinct types of potentiation: SDP, which persists for about 30 min; and LTP, which can persist for hours or longer. SDP is mimicked in hippocampal synapses by application of NMDA, an analogue of the excitatory neurotransmitter glutamate. Expression of SDP does not depend on protein kinase activity⁷. Induction of true LTP requires stimulation of presynaptic axons in addition to activation of NMDA receptors. The stimulation can cause release of an additional unknown neurotransmitter. In contrast to SDP, both induction and maintenance of LTP depend on the activity of protein kinases.

other proteases. Greenberg *et al.*²⁶ suggest that enhanced expression of genes encoding proteases such as calpains could increase the level of constitutive kinases in neurons, producing long-lasting changes in their properties. But there is as yet no direct evidence for involvement of calpain in LTP.

In their new work, Malinow *et al.*⁷ provide strong, if not airtight, support for the general idea that persistent protein-kinase activity underlies maintenance of LTP. These authors examine the effects on LTP of two protein-kinase inhibitors that diffuse into membranes and can act within cells. Sphingosine, a lipid precursor of gangliosides, prevents activation of both C-kinase and CaM kinase by binding to their regulatory domains; and H7, a synthetic organic inhibitor, competes with ATP at the active sites of protein kinases. When sphingosine or H7 is applied to hippocampal synapses during tetanic stimulation, induction of LTP is blocked, but sphingosine has no effect on the level of potentiation when it is applied during the maintenance phase. In contrast, application of H7 during the maintenance phase reduces synaptic strength to its preinduction level, and when it is washed away, synaptic strength rises again to the potentiated level. Thus, activation of one or more protein kinases appears necessary for induction of LTP, and a distinct constitutive protein kinase must remain active for expression of potentiation during the maintenance phase.

Unfortunately, both sphingosine and H7 are insufficiently specific to distinguish reliably among the several known protein kinases. Nevertheless, these data reinforce the notion that protein kinases are central to the mechanisms underlying LTP. These data are compatible with

elements of all three hypothetical mechanisms: activation of C-kinase and/or CaM kinase may be necessary for induction or maintenance of LTP, and constitutive forms of either or both could be generated by the action of calpain.

Much remains to be explained. None of the three mechanisms addresses the unusual feedback from postsynaptic site to presynaptic terminal that occurs during induction. The latest work, by Dumuis *et al.*²⁷, shows that binding of NMDA to its receptor releases arachidonate into the extracellular space, which could be a transsynaptic messenger able to travel backwards to alter presynaptic functions. Other entirely new messenger systems may be awaiting discovery.

Interpretation of physiological experiments is impaired by our inadequate knowledge of the molecular structure and function of components of brain synapses such as vesicles, receptors and the postsynaptic density, and our equally inadequate knowledge of the subcellular organization of second-messenger systems. As our understanding of the molecular organization of synapses becomes more sophisticated, we will undoubtedly move from our present, somewhat naive, formulation: one messenger → one enzyme → one effect, to a more accurate conception of second-messenger systems as highly interdependent regulatory networks. □

1. Rich, A. *Diving into the Wreck: Poems 1971-1972* (Norton, New York, London, 1973).
2. Morris, R.G.M., Anderson, E., Lynch, G.S. & Baudry, M. *Nature* **319**, 774-776 (1986).
3. Bliss, T.V.P. & Lomo, T. *J. Physiol., Lond.* **232**, 331-356 (1973).
4. Bliss, T.V.P. & Lynch, M.A. in *Long-term Potentiation: from Biophysics to Behaviour* (eds Landfield, P.W. & Deadwyler, S.A.) 3-72 (Liss, New York, 1988).
5. Zola-Morgan, S., Squire, L.R. & Amaral, D.G. *J. Neurosci.* **6**, 2950-2967 (1986).
6. Kauer, J.A., Malenka, R.C. & Nicoll, R.A. *Nature* **334**, 250-252 (1988).
7. Malinow, R., Madison, D.V. & Tsien, R.W. *Nature* **335**, 820-824 (1988).
8. Dolphin, A.C. *et al.* *Nature* **287**, 496-498 (1982).
9. Wilson, R.C. *J. Neurophysiol.* **46**, 324-338 (1981).
10. Malinow, R. & Miller, J.P. *Nature* **321**, 529-530 (1986).
11. Collingridge, G.L., Kehl, S.J. & McLennan, H. *J. Physiol., Lond.* **334**, 33-46 (1983).
12. Nowak, L. *et al.* *Nature* **307**, 462-465 (1984).
13. Kelso, S.R., Ganong, A.H. & Brown, T.H. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5326-5330 (1986).
14. MacDermott, A.B. *et al.* *Nature* **321**, 519-522 (1986).
15. Kaczmarek, L.K. & Levitan, I.B. (eds) *Neuromodulation: The Biochemical Control of Neuronal Excitability* (Oxford University Press, 1987).
16. Castagna, M. *et al.* *J. Biol. Chem.* **257**, 7847-7851 (1982).
17. Malenka, R.C., Madison, D.V. & Nicoll, R.A. *Nature* **319**, 774-776 (1986).
18. Akers, R.F., Lovinger, D.M., Colley, P.A., Linden, D.J. & Routtenberg, A. *Science* **231**, 587-589 (1986).
19. Lovinger, D.M. *et al.* *Brain Res.* **343**, 137-143 (1985).
20. Kennedy, M.B., Bennett, M.K. & Erond, N.E. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7357-7361 (1983).
21. Miller, S.G., Patton, B.L. & Kennedy, M.B. *Neuron* **1**, 593-604 (1988).
22. Lisman, J.E. & Goldring, M.A. *Proc. natn. Acad. Sci. U.S.A.* **85**, 5320-5324 (1988).
23. Lynch, G. & Baudry, M. *Science* **224**, 1057-1063 (1984).
24. Siman, R. & Noszek, J.C. *Neuron* **1**, 279-287 (1988).
25. Inoue, M., Kishimoto, A., Takai, Y. & Nishizuka, Y. *J. Biol. Chem.* **252**, 7610-7616 (1977).
26. Greenberg, S.M., Castellucci, V.F., Bayley, H. & Schwartz, J.H. *Nature* **329**, 62-64 (1987).
27. Dumuis, A. *et al.* *Nature* (in the press).

Mary B. Kennedy is in the Division of Biology, California Institute of Technology, Pasadena, California 91125, USA.

Daedalus

Heat rising

LAST week Daedalus decided that a plate of molecular sieve, hot on one face and cold on the other, would continuously absorb gas molecules onto its cold face and expel them from its hot one. It would be a molecular pump, capable of operating on any chosen gas. DREADCO's engineers, with their usual enthusiastic megalomania, are now building molecular sieve air-pumps as big and powerful as possible.

Their prototype pump is intended for air bearings, hovercraft and other air-cushion devices. A flat disk of an air-absorbing molecular sieve, with electric heating elements embedded in its lower surface, skims effortlessly over a plane floor on the expelled air blowing out beneath it. This simple, silent, elegant levitation system is a wonderful improvement on the noisy fans and blowers of current technology. Cheap and simple conveyor systems for factories and warehouses, and maybe even full-blown passenger-carrying hovercraft, should be possible.

But the DREADCO engineers want their pump to fly! As a first step, they replaced the wing panels of a conventional light aircraft with molecular sieve air-pump surfaces. The initial idea was to test the theory that the drag of a wing surface can be reduced by sucking in, or blowing away, the boundary layer that normally clings to it. But the engineers arranged the ferocious heating elements so that the top surface sucks air in, while the bottom one boils it out. During tests on the ground, the wings produced such a downdraught in their own right that the plane lurched vertically into the air without its pilot, turned over and crashed.

As a result of this accident, Daedalus is now devising the ultimate, simple, silent solid-state helicopter. It is little more than a big plate of molecular sieve with a web of burners on the lower surface. Their heat forcefully impels air downwards through the plate, creating a downdraught which provides both lift for the craft and air for combustion. If the molecular sieve material is chosen to absorb oxygen preferentially to nitrogen, the combustion is even more vigorous, heating the underside of the plate red-hot. The downflow of hot expanding gas forces the craft into the sky. Daedalus hopes to steer and propel it simply by tilting it sideways slightly, like a helicopter rotor; but problems of control, and where to put the pilot and payload, are still being debated. Elegant, reliable, infinitely manoeuvrable and with practically no moving parts, the new craft should revolutionize air transport. Already the speeding glowing disks in the sky above DREADCO's headquarters are giving rise to strange rumours. . . .

David Jones

Are racehorses becoming faster?

SIR—"Why aren't horses faster?" asks W. G. Hill in his comment¹ on the study by Gaffney and Cunningham² of a genetic trend in thoroughbred horse-racing performance. We suggest here that today's horses are faster and still improving, if only marginally. Figure 1 in the paper by Gaffney and Cunningham shows that between about 1840 and 1980, English thoroughbred horses reduced their winning times for the St Leger, Oaks and Derby races by about 12, 20 and 18 seconds, respectively. The resultant per generation improvements in winning times, assuming an average generation interval of 10.1 years³, are 0.4 to 0.8 per cent per generation. The Derby was won in 1988 by Kahyasi, whose time of 2:33.84 bettered by 0.06 seconds the record time set in 1987 by Reference Point: a one-year reduction of 0.04 per cent (again, 0.4 per cent per generation). Such improvements are modest when compared with the 1 to 3 per cent gain per year estimated for traits of economic importance in other species of livestock⁴. Nevertheless the horses are running faster.

Several factors might account for the lower apparent gain in response to selection among thoroughbred horses, in comparison with other categories of livestock subjected to breeding regimens. First, compared with factors such as weight gain, carcass composition, milk yield or fleece thickness selected for in pigs, cattle or sheep, race performance in horses involves such diverse components as muscle mass, relative lengths of limb segments, joint strength and stability, and aerobic capacity. Gains in some of these

might even be antagonistic to progress in others. Second, improvements in many domesticated breeds have emphasized characteristics very different from those presumably favoured in wild populations; under such circumstances, the initial gains from artificial selection can be quite rapid. In contrast, improvement in the running speed of horses extends a trend that goes back at least 50 million years to the Eocene; it is likely that current gains are occurring along the distant reaches of an asymptotic curve, where continued selection might be expected to produce relatively less net change. Horse breeders can take comfort from the realization that, however slight their gains might seem per generation, on the average they are hundreds of thousands of times more rapid than examples of vertebrate evolution documented from the fossil record⁵. In fact, to a palaeontologist the morphological correlates of the changes in modern horses might even appear punctuational.

ROBERT B. ECKHARDT

*Department of Anthropology and
Graduate Program in Genetics,*

DAVID A. ECKHARDT

JONATHAN T. ECKHARDT

*Department of Computer Science,
The Pennsylvania State University,
University Park,
Pennsylvania 16802,
USA*

1. Hill, W.G. *Nature* **332**, 678 (1988).

2. Gaffney, B. & Cunningham, E.P. *Nature* **332**, 722-724 (1988).

3. Mahon, G.A.T. thesis, Dublin Univ. (1980).

4. Smith, C. *R&D Agric.* **1**, 79-85 (1984).

5. Gingerich, P.D. *Science* **222**, 159-161 (1983).

An alternative view of the size of solar cycle 22

SIR—Several predictions of the size of the present sunspot cycle have been made¹⁻⁵ over the past year. They all suggest that the present sunspot cycle (cycle 22) will be larger than average, possibly of record size (larger than cycle 19 which had a maximum amplitude of 201.3, in terms of the moving average or smoothed sunspot number) or of near-record size (larger than cycles 21 or 18, the second and third largest cycles, respectively, in the modern era of sunspot observations).

A common thread in the predictions is that they are based on linear fits between the maximum amplitude of the sunspot cycle and the level of geomagnetic activity near cycle onset. Kane¹, for example, makes use of the minimum value of the aa geomagnetic index near the sunspot minimum; Gonzalez and Schatten² use separately minimum values of either the aa index or the Ap index; Thompson³ uses the number of 'disturbed days per year' (based on the Ap index); and Brown⁴ uses

the number of geomagnetic 'abnormal quiet days'.

Although the linear fits used to make the predictions have a correlation coefficient $r \geq 0.8$, they often have a rather large standard error of estimate (≥ 20), implying that the predictions may not be very reliable. In fact, a comparison of observed and expected values of maximum amplitude for modern era sunspot cycles, based on these single-variate fits, shows that they have differed by as much as 44 units of sunspot number.

In contrast to the above predictions are the so-called 'bivariate' fits^{6,7}, each based on the combined effects of a geomagnetic index and the level of sunspot number at cycle onset. Such bivariate fits have been shown to be highly correlative ($r \geq 0.95$) and to have a small standard error of estimate (≤ 10), implying that predictions of maximum amplitude based on the bivariate fits are inherently more reliable. Application of selected bivariate fits to

cycle 22 suggests that it will have a maximum amplitude larger than the mean maximum amplitude ($=116.3$, based on cycles 10-21), but one that is not of record size.

Of several bivariate fits, the most reliable for predicting maximum amplitude, $R(\max)$, is one that uses the minimum annual averages of sunspot number, $R(\min)$, and the Ap index, Ap (min), having a correlation coefficient of 0.997 and a standard error of only 3.9. For cycle 22, $R(\min)$ occurred in 1986, having a value of 13.4, and Ap(min) occurred in 1987, having a value of 11.0, implying that $R(\max)$ should be about 144.6 ± 7.8 (a 2- σ prediction interval).

Based on cycles 17-21, this particular fit has never erred by more than 4.1 units of sunspot number. Providing that cycle 22 is statistically no different from that of cycles 17-21, one infers that cycle 22 should be smaller than both cycles 21 (164.5) and 18 (151.8) and probably about the size of cycle 11 (140.5), making it the fourth or fifth largest sunspot cycle of the modern record. The value 144.6, a smoothed sunspot-number value, is equivalent to 139.4 when expressed as an annual average sunspot number. This latter value is near to just outside the lower limit reported by Brown⁵ for the maximum amplitude of cycle 22 (175 ± 35 , in terms of the annual average). Thus, cycle 22 probably will not be an exceptionally large, record-setting sunspot cycle, although it will be above average in size.

ROBERT M. WILSON

*Space Science Laboratory,
NASA Marshall Space Flight Center,
Huntsville,
Alabama 35812,
USA*

1. Kane, R.P. *Sol. Phys.* **108**, 415-416 (1987).

2. Schatten, K.H. & Sofia, S. *Geophys. Res. Lett.* **14**, 632-635 (1987).

3. Gonzalez, G. & Schatten, K.H. *Sol. Phys.* **114**, 189-192 (1987).

4. Thompson, R.J. *Austr. Phys.* **25**, 17-19 (1988).

5. Brown, G.M. *Nature* **333**, 121-122 (1988).

6. Sargent, H.H. III 28th IEEE Vehic. Tech. Conf. 490-496 (IEEE, New York, 1978).

7. Wilson, R.M. *Geophys. Res. Lett.* **15**, 125-128 (1988).

'Maize' in Somnathpur, an Indian mediaeval temple

SIR—Although Johannessen claims (*Nature* **332**, 587; 1988) that the object in the hands of the goddess in Somnathpur temple represents a maize ear, there is no evidence of maize cultivation at the time of construction of the temple in the thirteenth century AD, either in Karnataka or elsewhere in peninsular India. Maize cultivation in Karnataka commenced in the mid-1960s.

The similarity of the object described by Johannessen to a maize ear was first brought to our attention by Professor D.B. Lawrence (Minnesota University) in



Fig. 1 Male deity in lotus posture (*padmasana*) holding an object the upper half of which is beaded and the lower half smooth (arrow).

July 1970. A careful analysis of more than 50 friezes on the outer side of the temple led us to conclude that these objects do not represent maize ears (J.K.S. Sachan, *et al.* in *Advances in Crop Improvement* (eds R.B. Singh *et al.*) 41–48; Kalyani, 1982).

The Kesava temple of Somnathpur was built in 1268 AD by a high officer Soma or Somnatha under the Hoysala King Narsimha the third. The friezes were decorated by the chief sculptor Mallithamma. Both gods and goddesses are seen to hold an object having a fully or partially beaded type of ornamentation in one hand and *kalash* (pitcher) in the other (Figs 1,2). The objects are oblong, broadly cylindrical, and either conical or resembling a mango fruit in shape. In some friezes, the male deity is holding an object with only partially beaded ornamentation, that is, the upper headed half has a greater diameter than the lower smooth half (Fig. 1). In some sculptures the object is thicker in the middle and tapers at both ends. When some of these objects are viewed in isolation, a few of them may appear to resemble a maize ear in a gross way because of the beaded type of ornamentation (Fig. 2). But when all the sculptures are analysed, it becomes evident that the objects do not represent a maize ear.

The Hoysala tradition in sculpture is famous for its heavily ornamented images. In the friezes the beaded ornamentation is represented by the gold necklace, and also in the headgear, in *kardhani* worn around the waist and which hangs on the thighs, and *kalash*. Based on the evidence presented above, we can



Fig. 2 Female deity holding a beaded type of ornamentation tapering at both ends, claimed to be a maize ear.

exclude the possibility that the beaded object in the hand of the goddess is a maize ear.

M.M. PAYAK

Division of Mycology and Plant Pathology,
J.K.S. SACHAN

Division of Genetics,
Indian Agricultural Research
Institute, B-12,
New Delhi 110012,
India

Short but faithful pieces of ancient DNA

SIR—Pääbo and Wilson¹ point out the advantage of using *in vitro* DNA amplification by polymerase chain reaction (PCR)², rather than traditional cloning, in the analysis of ancient DNA from mummified tissues. We would like to add some data that arose from a series of experiments on a sample of Precolumbian maize. Although, as far as we know, the sample has not been submitted to mummification treatment of any sort, the data from it closely support Pääbo and Wilson's observations in showing both that only short pieces of DNA can be amplified from ancient DNA and that ancient DNA is faithfully amplified by *Taq* polymerase.

The sample comprised nine well preserved maize ears from a Huari tomb (Peru coast) kindly supplied by Centro Studi Ricerche Ligabue of Venice. Maize grains were radiocarbon dated by Geochron Laboratories MA to 980 ± 95 yr before present (reference year 1950).

To isolate nucleic acids, single maize

seeds were first washed with an SDS-EDTA-based mixture to remove possible external contaminants, soaked overnight in a medium of the same composition and then homogenized by mortar and pestle. Nucleic acids were extracted by phenol, phenol chloroform and chloroform, followed by ethanol precipitation.

After further purification by electrophoresis on low-gelling agarose and re-extraction, the DNA fraction was submitted to enzymatic amplification using an automatic thermal cycler³ and four PCR systems designed as follows. *CoxI* 'long' was aimed at the nucleotide sequence from -180 to -30 of the mitochondrial cytochrome *c* oxidase subunit I from fertile maize⁴; *CoxI* 'short' spanned a reduced portion of the same sequence (-160 to -30). Similarly, the *H2a* 'long' spanned the first 149 nucleotides of the clone *H2a* of a family of highly repeated sequences present in heterochromatin-rich maize⁵, while *H2a* 'short' covered only nucleotide 49 to 149 of the same sequence. The result of the amplification reactions (40 cycles) was that while the two 'long' PCR systems did not produce detectable amplification products, the two 'short' systems both generated amplification bands of the expected length (respectively 130 and 100 base pairs).

The 130-base-pair band produced by the 'short' *CoxI* system was further characterized by restriction digestion with *HaeIII*, terminal labelling with ³²P and fractionation on high-resolution polyacrylamide gel. This revealed the presence of a *HaeIII* restriction site located at the position expected from the corresponding nucleotide sequence of modern maize DNA — evidence that the ancient DNA had been faithfully amplified.

FRANCO ROLLO

AUGUSTO AMICI

ROBERTO SALVI

ANNAROSA GARBUGLIA

Dipartimento di Biologia Cellulare,
via F. Camerini 2,
I-62032 Camerino,
Italy

1. Pääbo, S. & Wilson, A. C. *Nature* **334**, 387–388 (1988).
2. Saiki, R. K. *et al.* *Science* **239**, 487–491 (1988).
3. Rollo, F., Amici, A. & Salvi, R. *Nucleic Acids Res.* **16**, 3105–3106 (1988).
4. Isaac, P. G., Jones, V. P. & Leaver, C. J. *EMBO J.* **7**, 1617–1623 (1988).
5. Viotti, A., Privitera, E., Sala, E. & Pogna, N. *Theor. appl. Genet.* **70**, 234–239 (1985).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of less than 500 words and five references. □

Out of the Garden of Eden

Alvin M. Weinberg

Nuclear Fear: A History of Images. By Spencer R. Weart. *Harvard University Press*: 1988. Pp. 535. \$29.50, £23.50. To be published in paperback next spring.

Collapse of an Industry: Nuclear Power and The Contradictions of U.S. Policy. By John L. Campbell. *Cornell University Press*: 1988. Pp. 231. Hbk \$32.50, £26.80; pbk \$11.95, £9.85.

CAN liberal democracy and nuclear energy coexist? This profound question perplexes Western democracies in the wake of Chernobyl, the Swedish rejection of nuclear energy, and the nuclear moratoria, *de facto* or *de jure*, in most of the Western world.

The answer hinges on two underlying conundrums. First, why have most people, particularly in the West, turned so strongly away from things nuclear? Secondly, why, in spite of public disaffection, does nuclear energy continue to do well in some liberal democracies, most notably France? Spencer Weart, a physicist and professional historian, addresses the first issue in *Nuclear Fear*; John L. Campbell, a political scientist, addresses the second in *Collapse of an Industry*.

What is there about nuclear energy that evokes so much public opposition? After all, hydroelectric dams, not to say coal-burning plants, have caused far more premature deaths than nuclear reactors; and Bhopal was in many ways much worse than Chernobyl. Yet the public accepts dams, coal and even noxious chemicals. Weart identifies "four main themes in the web of associations that influenced the way people think about nuclear power:

(1) the technical realities of reactors . . . from these realities . . . the hypothetical effects of widespread low-level radiation from a Maximum Credible Accident were selected and stressed";

(2) this biased selection resulted from "nuclear energy's social and political

associations", particularly its symbolization of all modern industrial society, with the concerns over the role of authority in modern civilization. But nuclear power became uniquely this symbol as a result of "(3) the old myths about pollution, cosmic secrets, mad scientists and apocalypse that were historically associated with atomic power and radiation, indestructible myths with deep psychological resonances." And finally

"(4) was the threat of nuclear war".

Nuclear Fear is then a remarkable historical examination of both these old myths and the newer social and political associations that led to nuclear energy becoming, according to Weart, the "condensed symbol of all of modern society". The book is not an ordinary history of the nuclear age like Richard Rhodes's *The Making of the Atomic Bomb*. It is rather a painstaking history of the imagery, the symbols, the associations connected with the nucleus — images that in final analysis determine popular attitudes towards nuclear energy. These images were usually created by the scientists themselves, for example by Frederick Soddy who in 1908 claimed that nuclear energy could make the world "one smiling Garden of Eden" (or I suppose my own characterization in 1972 of nuclear energy as a Faustian Bargain).

The images were then elaborated and popularized by authors, poets, journalists and television commentators. Which of the two visions — "White City" (conjured up by the Chicago 1893 World's Fair) or

the Hiroshima-like "Desert of Ash" — was embraced by the professional purveyors of images depended upon their political proclivities and even their psychoanalytical antecedents. At the deepest level, Weart identifies "White City" imagery with a social outlook that saw salvation in culture, that is in ever-advancing technology; by contrast the "Desert of Ash" imagists distrusted technology, and sought salvation in nature.

Nuclear Fear is an extraordinary *tour de force*: Weart's historical scholarship is overwhelming, his writing is sprightly and his main theme — that nuclear energy's currently bad image has deep and unique antecedents — reveals powerful new insights as to why so much of the public distrusts the nucleus. On the other hand, where Weart doffs his historian's hat and dons that of a psychoanalyst, complete with images of the womb and the phallus, I am slightly bemused. But these psychoanalytical forays are a small part of the book, and those of psychoanalytical bent can judge Weart's bona fides as a Freudian better than I.

Collapse of an Industry takes up where *Nuclear Fear* leaves off. Given that a large part of the public has become disaffected with nuclear energy, how can one explain the success of nuclear energy in France and its failure, at least for the time being, in the United States, Sweden and possibly West Germany. Campbell analyses in detail the situation in the United States. As befits a political scientist, he finds the failure to lie in the mismatch between the imperatives of nuclear technology, and the institutions in the United States responsible for forming and for implementing nuclear policy. He discusses four principal failures. First, lack of reactor standardization, which he correctly attributes to the extreme fragmentation of the nuclear industry in the United States. Secondly, insufficient attention to reactor safety, which he attributes, fundamentally, to the Atomic Energy Commission's conflicting role as both the promoter and

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

regulator of nuclear power. Thirdly, the difficulty in raising additional investment capital as costs escalated, especially in response to the obstructionism of the nuclear critics (for Campbell, direct government financing, as in France, would have been a better policy). And finally, the waste disposal impasse, which Campbell observes is exacerbated both by the failure of the private sector to reprocess fuel, and by American federalism which enables a state selected as host for a disposal site to claim its right to reject the repository.

Campbell tends to overstate the magnitude of the 'failures' of nuclear energy in the United States — after all, some 18.5 per cent of the nation's electricity is now produced from uranium — and this betokens a (possibly inadvertent) anti-nuclear tilt. Nevertheless, his attribution of the shortcomings of the nuclear industry in the United States to the fragmented, decentralized and open structure of the mechanisms for implementing policy has much truth in it. France, by contrast, is a mirror image; the nuclear industry is a state monopoly and both formation and implementation of policy are centralized, unified and closed. Thus France offers far fewer opportunities for outsiders to intervene either in the formation or in the implementation of nuclear policy than even Sweden and West Germany, and certainly the United States. This, in Campbell's view, as well as that of other observers such as J. Barkenbus, goes far to explain why France has succeeded, and the United States has not.

But this is surely only part of the story. The main reality, one that possibly transcends the difference in political structure, is the degree of need for nuclear power. France has little fossil fuel, and nuclear power met a more pressing need there than in the United States.

This is not to say that political customs and institutions are unimportant. Sweden has even less fossil fuel than France yet has rejected nuclear energy. Campbell explains the Swedish experience by the very openness of the Swedish system of forming policy. Sweden encouraged the broadest participation of its citizens in setting nuclear policy — the result, the end of nuclear power in Sweden. Thus, although Campbell (who seems to take neo-Marxism or at least anti-capitalism somewhat seriously) extols the virtue of open democracies in the early part of his book, by the time he reaches the end he resignedly admits that if the United States increases "democratic participation, particularly during policy formation, it is feasible that nuclear critics could put an official end" to nuclear power. In short, Campbell, the political scientist, all but says that nuclear power and open liberal democracy are incompatible! This is not surprising, given Weart's demonstra-

tion that nuclear fears are deep and are widely held.

Is there, then, no future for nuclear technology, at least in the most open liberal democracies? In a concluding personal note, Weart offers some hope on this score. He would exorcize the fearful imagery of nuclear disaster, and thus allow nuclear energy and democracy to coexist, primarily by changing the realities (as opposed to the myths) of the nucleus. As for nuclear energy, he urges developing nuclear reactors that are inherently safe, and can be seen to be so. As for nuclear war, he is attracted by Immanuel

Kant's comity of liberal democracies on the grounds that democratically elected governments never make war on one another. Although this advice is outside the scope of *Nuclear Fear*, it appeals strongly to me. I can only wish that Weart's sensible recommendation is taken seriously both by the nuclear establishment and by its critics. □

Alvin M. Weinberg, Institute for Energy Analysis, PO Box 117, Oak Ridge, Tennessee 37830, USA, was Director of Oak Ridge National Laboratory from 1955 to 1974. He was a member of the war-time nuclear laboratory in Chicago, and in 1945 was the proposer of the pressurized water reactor.

A to X of B cells

A. R. Williamson

B Lymphocytes in Human Disease. Edited by Graham Bird and Jane E. Calvert. Oxford University Press: 1988. Pp. 513. £55, \$98.

LYMPHOCYTE biology involves a *pas de deux* between B and T lymphocytes. But each of these types of cell has an independent development and its own varied pattern of functional differentiation.

The B cell, with receptor immunoglobins and the capacity to differentiate into an antibody secreting cell, is a more advanced evolutionary development than the T cell, which has receptors capable of recognizing non-self only when associated with self major histocompatibility molecules. Recent discoveries relating to T-cell receptors and MHC class I molecule structure have held the limelight, so it is perhaps timely that a new volume should be devoted to B lymphocytes. Moreover, the value of immunology to the clinician is being enhanced by our increased understanding of the immune system, and thus a presentation entitled *B Lymphocytes in Human Disease* should be doubly welcome. It is, however, hard to be very enthusiastic about this multi-author book.

One difficulty is in understanding who the intended readers are. Half of the book is devoted to the biology of B cells, starting with a chapter on antibody structure and function, and running through some good reviews of antibody genes (with a separate chapter on isotype switching) and most stages of B-cell development, differentiation, activation and signalling. Only the second half of the book is devoted directly to the subject matter promised by the title. Here the material seems suited to the clinician or to the immunologist moving to clinical interests. Each review can be read in isolation from the others and from the first half of the book (there are a few cross references but they aren't essential to understanding). Antibody deficiency states are catalogued and possible mecha-

nisms briefly reviewed. It is a pity that the recent and exciting evidence for the molecular defect in SCID mice appeared too late for inclusion in this chapter.

Reviews of B-cell neoplasia and of the underlying genetic mechanisms provide, respectively, an update on B-cell differentiation and on the relevant immunoglobulin genetics that do not require the reader to refer back to the first half of the book. In the chapters on mechanisms for auto-antibody formation, anti-idiotypes and idiomotype networks are discussed clearly enough by several authors, but without any new insight being revealed.

By far the best understood autoimmune disease, myasthenia gravis, is well reviewed in the last chapter (the only one devoted to a single disease). This chapter also includes experimental data, something almost totally absent elsewhere in the book. Other contributions have a few tables, but generally there are too few good diagrams.

The penchant of the immunologist for alphabet soup (from A to X), but not for pithy acronyms, is demonstrated by the six pages of abbreviations preceding the introduction. Soup and the number of cooks are easily invoked when describing a multi-author volume. But this one can be taken à la carte for the worth of certain courses, even though the whole menu is less digestible. □

A. R. Williamson was until recently Director of the Glaxo Institute for Molecular Biology, Route des Acacias 46, 1227 Carouge, Geneva, Switzerland.

New in paperback

- *Superstring Theory Vols 1 & 2* by M. B. Green, J. H. Schwarz and E. Witten, one of the series Monographs on Mathematical Physics. Publisher is Cambridge University Press, prices are Vol. 1, £12.50, \$19.95; Vol. 2, £15, \$24.95. For review see *Nature* **328**, 26 (1987).
- *Selling Science: How the Press Covers Science and Technology* by Dorothy Nelkin. Publisher is W. H. Freeman, price is \$9.95, £7.95. For review see *Nature* **328**, 677 (1987).
- *Other Worlds: Space, Superspace and the Quantum Universe* by Paul Davies. Publisher is Pelican, price is £3.99. For review see *Nature* **287**, 567 (1980).

Comprehensive colloids

A. D. Buckingham

Surface Forces. By B. V. Derjaguin, N. V. Churaev and V. M. Muller. *Consultants Bureau (Plenum): 1987. Pp. 440. \$114, £67.*

THE NAME of the distinguished Soviet scientist B. V. Derjaguin is well known throughout the world. Derjaguin began publishing in the field of surface effects in 1932, and has written many papers and books on the subject of colloid science. In 1941, with L. Landau, he published a classic paper on the theory of the stability of lyophobic colloids; in the occupied Netherlands E.J.W. Verwey and J.Th.G. Overbeek independently solved the same problem and the theory is universally denoted 'DLVO'.

Derjaguin has been particularly concerned with the influence of solids on the structure and properties of liquids in contact with them. The changes due to these 'surface forces' may extend over distances that are large compared to molecular dimensions but small in relation to macroscopic objects. The forces have many notable manifestations, including their effects on colloid stability, the properties of thin interlayers and films, and flotation, and implications in water treatment, soil science and microbiology.

This book is claimed to be the first comprehensive monograph devoted to surface forces. It contains eleven chapters, a brief conclusion and a short subject index. Emphasis is placed on general theory and on the application of surface

forces to thin films, forces between macroscopic objects, boundary layers in polar liquids, the stability of lyophobic colloids, wetting and osmosis. From the preface we learn that Derjaguin contributed to all eleven chapters, Churaev to five and Muller to four.

The book conveniently presents the principal contributions of the Moscow group to colloid science, giving a compact account of a large amount of research in an important area. The approach is generally through bulk properties, such as susceptibilities, rather than through a truly molecular description. Apart from a forward glimpse in the Conclusion, computer simulation is not mentioned. The monograph is not appropriate for students wishing to learn about the fundamentals of intermolecular forces — the permittivity enters the theory at an early stage — and there are some irritating deficiencies; for example neither f nor h in equation (1.4) is defined, and SI units are not used.

It is appropriate to compare the book with J.N. Israelachvili's *Intermolecular and Surface Forces with Applications to Colloidal and Biological Systems*, published by Academic Press in 1985. Israelachvili adopts a rather more molecular view and a tighter focus; his volume is of wide appeal and excellent for those beginning research in molecular biology. It is much the more attractively produced of the two, but it must be remembered that *Surface Forces* has been translated into English. The new generation of researchers will benefit greatly from reading both books. □

A. D. Buckingham is a Professor in the University Chemical Laboratory, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK.

Blood companions

Lucio Luzzatto

Atlas of Blood Cells: Function and Pathology, 2nd edn. Edited by D. Zucker-Franklin, M. F. Greaves, C. E. Grossi and A. M. Marmont. *Edi Ermes, Via Timavo 12, I-20124 Milan/Gustav Fischer, Stuttgart: 1988. Two volumes, pp. 777. £154.69, \$225.*

HAEMATOLOGISTS have long been producers and admirers of good pictures. In the first half of this century a haematology textbook owed its success as much to the painstaking work of unnamed artists, who spent hours in front of the *camera lucida* of a microscope, as to the text written by its author. When hand-made drawings yielded to photomicrographs, the consequent greater objectivity was at first associated with a sharp drop in quality. Now photomicrography has come of age,

and when combined with the resolving power of transmission electron microscopy, the three-dimensional plasticity of scanning electron microscopy, the specificity of cytochemistry and of monoclonal antibodies, and with the real beauty of immunofluorescence, the results can be quite stunning. This two-volume atlas is a spectacular example of how the potential of various imaging techniques can be harnessed to help understand the structure and the biology of blood cells.

The subject is so vast that the authors must have been confronted with hard choices. A classical distinction in studying diseases has been between those that are congenital and those that are acquired. The approach to elucidating mendelian genetic disorders has capitalized on the notion that there must be a single lesion, and the task was to pinpoint it. Haematology has been a prime beneficiary of this approach, as haemoglobinopathies have become the prototype of molecular



Cetacean illustration — Whaling at Goro, Hizen Province by Hiroshige, a 19th-century master of the Japanese print. The picture is a detail of a colour original in *Whales* by Jacques Cousteau and Yves Paccalet, a beautifully illustrated large-format volume published by W. H. Allen/Planet, price £25; in the United States available from Harry N. Abrams, \$49.50.

pathology. Why has the understanding of acquired blood disorders, such as leukaemia, not reached this stage yet? There are at least two fairly obvious reasons. First, in spite of Boveri's early idea that something is wrong in the chromosomes of tumour cells, it has taken too long to realize that specific changes in the genome of somatic cells underlie many acquired diseases, just as specific changes in the genomes of germ cells underlie inherited diseases. Secondly, the somatic genomic lesions leading to leukaemia form a stepwise sequence, rather than being a single lesion. Thus, although in their preface the authors have coined the phrase "molecular morphology", morphology remains more important where knowledge at the molecular level is less advanced.

In the *Atlas*, the emphasis is therefore rightly placed on leukaemias and lymphomas. In this area they constitute much more than an atlas because they do an outstanding, up-to-date job of showing how gene rearrangement analysis of DNA, conventional phenotypic markers and the remarkable insights of traditional morphology come together in working out the taxonomy of these disorders. Also in this area, there is an enriching and authoritative chapter by Janet Rowley on chromosomal abnormalities. At the same time it is refreshing to see much space devoted to less-fashionable topics, such as

the fine structure of eosinophils and basophils.

In a work of this magnitude, there are bound to be points to quibble with. Most of the illustrations are superb, but on p.160 two myeloblasts are compared at different magnifications, and one of them is out of focus (no wonder the other appears more mature!). Some of the pictures of malaria parasites are also sub-standard. As for omissions, it is a pity not to find an illustrated discussion of the erythrocyte cytoskeleton in a book that is otherwise an excellent introduction to cell biology. Also, an atlas is made of maps, and although those of the immunoglobulin genes are generously portrayed, one misses those of thalassaemia mutations and of the chromosomal localization of genes expressed in blood, other than oncogenes.

For the past 20 years, experimental haematology has been largely identified with culturing haemopoietic cells. Much work has been frustrated by the extreme difficulty in isolating and obtaining the required growth factors, but with the cloning of the respective genes it is now possible, for instance, to grow erythroid cell colonies in a fully synthetic medium containing erythropoietin produced in an expression vector. The first chapter of the *Atlas* provides a good introduction to perceiving how these culture systems will help the understanding of the cellular and humoral control of haemopoiesis. The *in vivo* counterpart of this approach, at the clinical end of the spectrum, is illustrated in the last chapter; the senior author, Dan Thomas, is the justly acclaimed pioneer of bone marrow transplantation (BMT), the single most significant advance in haematological therapy. Reconstitution of haemopoiesis after BMT is also a dramatic *in vivo* experiment in developmental biology, which may tell us something about the elusive stem cells and perhaps give us a way to isolate them. These are the cells that will need to be handled if gene therapy experiments aiming to correct blood disorders are to succeed.

For the practising haematologist the *Atlas* will be a well-deserved luxury; for the scientist who wants to know about the potential of blood cells to investigate cell biology it is a perfect complement to *The Molecular Basis of Blood Diseases* edited by Stamatoyannopoulos *et al.* (reviewed in *Nature* 329, 210; 1987); for every laboratory of diagnostic or research haematology it is a must. But I hope the next edition will include a picture at atomic resolution of how erythropoietin fits the erythropoietin receptor: that will be molecular morphology at its best. □

Lucio Luzzatto is Professor in the Department of Haematology, Royal Postgraduate Medical School, Hammersmith Hospital, Du Cane Road, London W12 0HS, UK.

Mental manoeuvres

Donald Broadbent

The Making of Cognitive Science: Essays in Honor of George A. Miller. Edited by William Hirst. Cambridge University Press: 1988. Pp. 284. £25, \$29.95.

GEORGE Miller belongs to the small minority of key figures who have altered the shape of a subject. By origin a psychologist, he seized on the possible connections between that discipline, mathematics, linguistics and engineering, and showed that it was possible to discuss 'mental' events in precise and scientific terms. If it is now sensible to speak of a single area of 'cognitive science', as an offspring of the original disciplines, George Miller has a good claim to pater-



From *The Making of Cognitive Science*

George Miller — startling ability.

nity. Unusually for a *Festschrift*, the tributes to him and the acknowledgements of his work to be found in the book are not exaggerations but are sober truth.

In plan, the book follows his career and samples his main areas of research. The first section contains chapters about his early work in the Harvard Psychoacoustic Laboratory, each discussing a technical area clearly and intelligibly enough for the non-specialist to realize why Miller is of such standing. The second section deals with the later founding of the Center for Cognitive Studies, still at Harvard, and reports the experience of people there rather than considering the problems they worked on. The impact of the Center was to create a group dedicated to examining 'mental' events, with implications far outside Miller's immediate areas of interest. As just one example, it was the first university laboratory to use a computer to control psychological experiments.

These two sections are probably the strongest in the book. Generations of students know Miller through his elementary text, a warm and humane volume, centred on personal interactions and contributions, and it may be salutary for them to be reminded of his hard mathematical and experimental base. It was his emphasis on rigorous and precise formalization, and not simply his broader human sympathies,

that made him able to appreciate the importance for other disciplines of Noam Chomsky's mathematicization of grammar. He thus launched a whole generation of psychologists on experimental studies, largely aimed at establishing whether human use of language could be modelled by Chomsky's formulation. That enterprise is described in two of the contributions in the third section. The book makes it clear that the enterprise failed, in the sense that this form of grammar is a poor theory of human performance at processing sentences. Yet experiment and theory about such processing, with the numerous complex internal operations it requires, were shown to be possible. They will never now disappear from psychology.

Part of the reason for the inadequacy of early syntactic theory to produce a theory of human language use lies in the pervasive role of our knowledge about the world. In Quillian's example, that knowledge is revealed by the assignment of different syntactic structures to the sentences "He had some dirt from his farm by the Hudson river" and "He had some dirt from his farm by his foot". That is why, on moving from Harvard to Rockefeller in 1968, Miller increasingly turned to semantics and continued to do so at Princeton after 1979. The fourth section contains chapters that give suggestive examples of topics that can be studied in this area, and a hope for the future in the procedural approach urged by Johnson-Laird. This research area, however, is not as advanced as the earlier ones described in the book, and that is even more true of the attempts to link brain damage and function in the fifth section. The two chapters concerned return to accounts of personal experiences of Miller's startling ability to stimulate those around him, as demonstrated in his Rockefeller years; but the examples of neurological cases are largely from other laboratories. Finally, the sixth section examines the current status of cognitive science, through the eyes of a philosophical colleague at Princeton.

Hardly anybody will find all of the contributions suitable reading. For the enquiring outsider, some parts make a good introduction to the value of Miller's work. The general reader will find others too specialized and without a close connection to Miller's own thought, even though they report excellent technical work. For those who have known him, other parts again will be pleasant reminders of his lively and engaging personality. But the best monument to Miller is not simply this volume; like Wren's, it lies in everything we see around us in the study of cognition. □

Donald Broadbent, Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK, is a member of the external staff of the Medical Research Council and chairman of the ESRC-MRC-SERC Initiative on Cognitive Science and Human-Computer Interaction.

Dissipation and noise immunity in computation and communication

Rolf Landauer

IBM Research Division, T. J. Watson Research Center, Yorktown Heights, New York 10598, USA

Reversible computers which carry out each step without discarding information can, in principle, dissipate arbitrarily small amounts of energy per step if the computation is carried out sufficiently slowly. This has caused a re-examination of energy requirements in communication and measurement. There also, it is only those steps that discard information which have a lower limit on energy consumption. Such steps can be avoided in the transmission of information.

OVER recent decades, energy required to perform an elementary logic operation has decreased rapidly (Fig. 1; adapted from ref. 1). One is led to ask whether the laws of physics set a fundamental lower limit to this energy. Such limits occur in other fields; for example, the second law of thermodynamics was found to set a limit on the efficiency of thermal engines, and did so in a way that anticipated all possible further inventions and refinements. Similarly, Shannon's channel capacity theory² sets limits on the information transmission rate of a noisy communications channel, which cannot be circumvented by clever modulation and demodulation schemes. Questions of this sort in regard to computational energy requirements were discussed from the beginning of the modern digital computer. One can simplistically associate the typical thermal energy kT of a degree of freedom with a binary variable, or a logic gate. 'Associate' is, however, an ambiguous term, and it takes a leap of faith to equate this to a requirement for energy dissipation per elementary step. These early discussions are typified in Brillouin's well-known book³, which, despite a chapter entitled *The Problem of Computing*, avoids confronting the details of the computational process. The book does not discuss the physics behind simple logic operations such as AND and OR; much less does it discuss a total system such as a Turing machine or a cellular automaton. Here I shall describe the progress toward more definitive answers.

There are three closely related areas in which we can discuss energy requirements, specifically the computational process, the measurement process and the communications channel. These areas are related but not equivalent. Figure 2 stresses the distinction, for example, between computation and communication. We will first discuss computation. We will apply the insights gained there to communication and, to a lesser extent, measurement. Modern studies of the communications channel date back to Shannon². Questions about the energy needed in measurement began with Maxwell's demon⁴. If, in discussing the newer aspect—computation—I am led to imply that the older fields did not get it quite right, the assertion is bound to be seen as controversial. I do not expect instant agreement with such views, but only that they will be given a conscientious hearing.

Computation

Figure 3 illustrates symbolically the phase space of a computer. When we discard information, two different initial states, A and B, are mapped into the same final state. Information is discarded continually in normal computer operation. It takes place when we erase a bit in a memory, and also in typical logic operations, such as AND or OR, that have two or more binary inputs but only a single output. Consequently, the output cannot contain as much information as the input. In a conservative system,

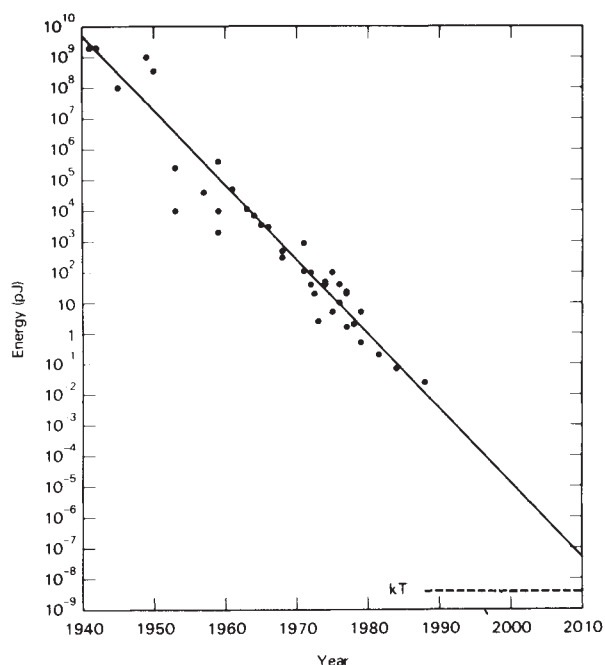


Fig. 1 The decrease in energy dissipated per logic operation over recent decades.

without friction, we cannot compress phase space in the way shown in Fig. 3; discarding information requires dissipative processes. For a more specific example, let a ball on the ground represent a '0' and a ball held in the hand, above the ground, a '1'. Can we map both of these into the '0' state, thus erasing the bit? If we simply let go of the ball in the hand, it will continue to bounce, in the absence of friction, and will not end up in the '0' state, but in the presence of friction it is clear that both initial states may end up in the '0' state. We can make the same point by allusion to adiabatic demagnetization, a well-known low-temperature refrigeration process. In this process a set of spins are aligned with a magnetic field. When the field is taken away the spins randomize their orientation and the resulting entropy increase is taken from the environment, cooling that environment. When we erase information we do the opposite, going from a diversity of states toward a standardized one. It can be expected that in this case the heat flow is then towards the environment.

There are many ways to encode information in a computer such that the energy of a '1' and a '0' are identical—for example, by magnetizing a volume in one direction or the opposite, or

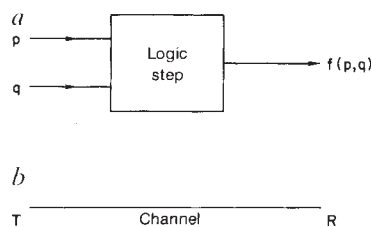


Fig. 2 Information processing and transmission. Logic step in *a* uses a nonlinear interaction between two binary inputs *p* and *q*. The transmission channel in *b* hopefully reproduces its transmitted input (*T*) at the receiving end (*R*).

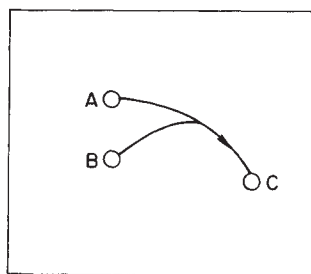


Fig. 3 Symbolic sketch of phase space of a computer, showing information loss in the transition from *A* or *B* to *C*.

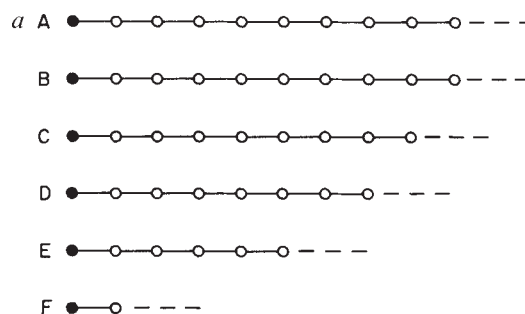


Fig. 4 *a*, One-to-one computation. The left-hand end of a horizontal chain represents the initial state, and forward computation represents motion to the right, through a sequence of states represented by successive open circles. Different letters correspond to different initial states, that is, different programs. *b*, Information-discarding junction. Two computational paths, moving to the right, merge into one.

by having a superconducting loop current flow in one direction or the other. The energy of the computer is then independent of its information content. Figure 4*a* shows a sequence of computational trajectories in such a system, for which no information is discarded and computation proceeds through a sequence of one-to-one mappings. In this system we avoid processes of the type shown in Fig. 4*b*. Computation can, in fact, be performed using such a sequence of 1:1 mappings^{5,6}. But noise can cause this computational sequence to move backwards, and, in the absence of a net biasing force pushing in the desired forward direction, the computation will actually move diffusively. To obtain predictable forward motion in addition to the purely diffusive motion, we must apply a force that pushes the computation to the right, in much the same way as an electric field causes an electron in a solid to move with a drift velocity. For small bias forces the electron (or computation) will move slowly, with both a velocity and a dissipation per step that is proportional to the applied force. Thus, if we are willing to compute slowly, we can do so with as little energy expenditure per step as desired. Note that the noise need not be the minimum thermal equilibrium noise from the system's finite temperature *T*, but can be larger, including additional, more practical sources of noise. Note also that a single step in Fig. 4*a* represents the forward evolution of the whole computer, not just of a single logic gate. Our discussion of Fig. 4*a* assumes that the computation is carried out with machinery which, as in the case of electrical current or of fluid flow, exhibits frictional forces proportional to velocity, rather than static friction.

The most common logic functions are not 1:1 mappings, but they can be imbedded in larger functions which are 1:1. Typically, this results in the production of extra outputs, which are not needed in the computation but which cannot be discarded if unnecessary dissipation is to be avoided. At the end of the computation we can copy the desired results from the output register. (As we show subsequently, copying can also be done with as little dissipation as desired, if done sufficiently slowly.) But during the computation, we will have accumulated many irrelevant but unavoidable outputs. What do we do with these? After all, we cannot just keep using more and more storage space as we go on to other computations. Bennett's unantici-

pated insight^{6,7} to cope with these outputs involves unwinding the program, that is, running it backwards. In view of the 1:1 nature of all the steps, this can easily be achieved. One possibility is to apply a small bias force to the computational trajectories of Fig. 4*a*, so as to drive the computation back to its initial point. This will clear out the unwanted information accumulated during forward computation. When, on the return trip, we arrive back at the original initial state, we can erase that information. To do so requires dissipation, albeit dissipation proportional to the size of the initial program and data, which can be much less than the number of steps in the computation or the amount of undesired information at the end of a computation. Alternatively, when we return to the initial state, and if we still have a copy of that initial state elsewhere, we can 'uncopy' the machine's state. Just as copying can be done with arbitrarily little dissipation, we can carry out the inverse operation with equally little dissipation. Uncopying differs from erasure; we return to this point later.

Figure 4*b* shows what happens when information is discarded: two tracks merge to form a single subsequent computational trajectory. Assume for simplicity that the two incoming tracks on the left are equally likely. Without bias force, and in thermal equilibrium, the phase space to the left of the junction is larger than that on the right. The system will prefer to stay to the left, unless it is driven predictably to the right by a potential drop $kT \ln(2)$, which is adequate to offset the entropy drop favouring the left-hand side. Thus, an energy $kT \ln(2)$ per erased bit has to be dissipated. We could rely on fluctuations to push us in the desired direction, past the junction in Fig. 4*b*, but a sequence of successive events of this type would then make passage past many such junctions an exponentially unlikely event.

Figure 4*a* is symbolic; it does not specify the kinetics which take us from one state of the computer to the next. The literature contains about half a dozen different suggestions (summarized in ref. 8) for how to achieve so-called reversible computation. We will allude, briefly, to two of these.

Figure 5*a* illustrates a three-input/three-output logic function, which I have called the Fredkin gate, after Fredkin¹⁰, who first perceived the importance of this function. With this gate, all the essential functions of a computer can be carried out¹⁰. Figure

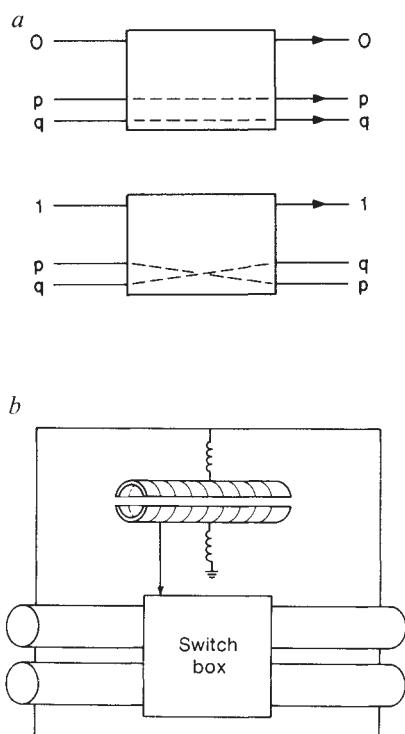


Fig. 5 *a*, A Fredkin gate. This has three inputs, shown on the left, and three outputs. The uppermost input is transmitted unchanged and controls whether the two lower tracks are also transmitted unchanged, or caused to cross over. This logic function loses no information and is its own inverse. *b*, A physical realization of the Fredkin gate. A ball moving along the top channel controls the gate and arrives at the split pipe, shown at the top, before the balls along the two bottom channels arrive at the switch box. A ball entering the top split pipe pushes the two halves apart and, in doing so, does work against the springs. This energy is retrieved, however, after the ball leaves the split pipe at the right-hand end. The springs serve to ensure that the split pipe remains closed if a zero signal (no ball) enters. The split pipe is coupled, in turn, to the switch box by a hard linkage, which controls the internal structure in the switch box so as to give the switching action shown in *a*.

5*b* shows a particular physical embodiment of this function, in which information is carried by balls moving along pipes or other guiding machinery, and where information is denoted by the absence or presence of a ball. The springs in Fig. 5*b* are needed to ensure that the switch remains in the uncrossed position if there is no ball in the split pipe. The reliability of the computation depends on the probability of thermal errors, and thus on the energy required for random thermal spring compression. By making the springs sufficiently stiff we can obtain any desired immunity to thermal errors. The energy needed to compress the springs intentionally is, of course, not dissipated, but is recaptured when the balls leave the split-pipe section. As mentioned already, we assume that the frictional forces for the machinery in Fig. 5*b* are proportional to velocity. A great deal more needs to be said^{11,12} to demonstrate how a complete reversible computer can be synthesized from this mechanism. For example, consider the case where the top left-hand input in Fig. 5*a* is represented by r , the middle input is s and the bottom line input is kept fixed at 0. The lower right-hand side output is then rs (the logical 'and' of r and s , which is 1 if and only if both r and s are 1), and the middle r.h.s. output is $s\bar{r}$ (the logical 'and' of s and of the negation of r). Figure 5*b* is clearly a symmetric device; its function can be inverted by simply running the balls back through the device.

Another reversible computer scheme invokes a particle in a time-varying potential well (Fig. 6). Such time dependence can

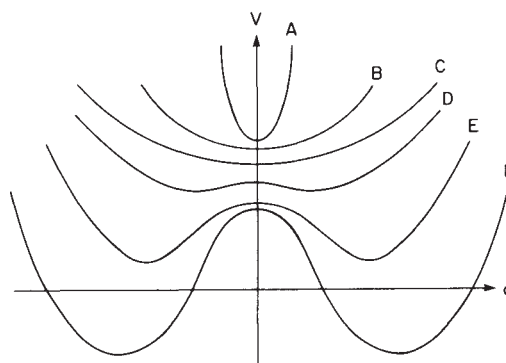


Fig. 6 Time-dependent potential well, which goes from a single minimum at A to a deeply bistable state at F, and then returns to A. Relative vertical displacement of curves is selected for clarity only. The displacement coordinate is represented by q .

be produced, for example, by cyclically bringing electrical charges towards and away from a charged information-bearing particle, whose motion is limited to a one-dimensional track. In this modulated potential scheme⁸ an odd number of deeply bistable wells are coupled to other wells that are just undergoing bifurcation. The bifurcating well will therefore be pushed into one or the other of its two developing valleys, depending on the state of the wells exerting the driving. The driving wells will be restored subsequently to their monostable state (top curve of Fig. 6) when they will again be ready to receive new information. This is an adaptation of a scheme proposed more than three decades ago by von Neumann¹³ and by Goto¹⁴ for computation using sub-harmonic parametric excitation. This general concept, using devices which are cycled back and forth between a monostable state and a bistable one, has since been adapted to other technologies, for example Josephson junction circuits^{15,16}. This application is particularly significant, because it is a genuine example (in possible contrast to the mechanical examples we have used in this discussion) of a system that has frictional forces proportional to velocity.

In these schemes we use majority logic to do computation. There is an odd number of inputs (typically three) to a given well. The well under consideration then gets set according to the majority vote of the stages exerting an influence on it. If, for example, we have one input set at zero, with the other two inputs being variable, we get a '1' output if and only if both of the variable inputs are '1'. In this scheme we cannot perform a negation directly. Therefore, we need dual rail logic, in which we carry along both a variable and its complement, and perform negation by interchanging the two.

If the time changes shown in Fig. 6 are slow enough, the probability distribution for the particles will remain close to the Boltzmann distribution. If the coupling forces (the springs in Fig. 7) that push the particle to one side or the other are large enough, then this Boltzmann distribution can be used to give any desired reliability. This is to say, the error probability resulting from a particle ending up on the wrong side can be made as small as desired. Furthermore, if we move the information-bearing particle sufficiently slowly, its frictional losses can be made arbitrarily small: there is no dissipation from the time dependence of the potential in Fig. 6 if the particle does not move. Incidentally, this scheme is not (at least in its simplest form) strictly an embodiment of Fig. 4*a*. Rather, the externally applied time modulation 'paces' the computation; it does not move diffusively. This is analogous to the controlled motion of a particle along a track in Fig. 4*a*, produced by trapping the particle in a potential well and then moving the well at a predetermined speed¹⁷.

Figure 7 shows the simplest sort of information transfer possible under this scheme. The springs shown in Fig. 7 are somewhat

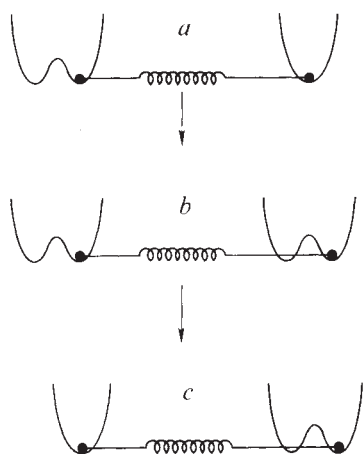


Fig. 7 Coupled particles in time-dependent potential wells. In the transition from *a* to *b*, the information in the bistable well on the left sets the state of the one on the right. In the transition from *b* to *c* the well on the left is restored to a monostable state and is ready to receive new information.

symbolic. They represent coupling devices which tend to bring particles in the coupled wells to positions which deviate equally from the middle of their respective potential. If all the wells were in a straight line, and equidistant, we could use 'real' springs for this purpose. In a computer, however, we will want to couple wells of various separations which are not necessarily parallel to the direction of particle motion. This will require springs which are tailored to their particular set of wells, or else more complex mechanical linkages to serve our intended purpose. In Fig. 7*a* the information in the well on the left is about to be transferred to the well on the right; this process is completed in Fig. 7*b*. We can then restore the left-hand well to its monostable state (Fig. 7*c*), where it is ready to receive new information. The copying, or transfer, in the transition from Fig. 7*a* to Fig. 7*b* is analogous to the information transfer which is part of a measurement cycle. Figure 7 demonstrates that, in contradiction to much of the literature in the field, this is not the part of the cycle that requires minimal dissipation. As argued elsewhere^{5,7,18,19}, dissipation arises from the subsequent resetting of the meter after it is decoupled from the object that is measured. The transition from Fig. 7*b* to 7*c* can be considered as the inverse of the copying process: one of two bits, guaranteed to be identical, can be restored to a standardized state. This is distinct from erasure, which would consist of a restoration of the left-hand well in Fig. 7*b* to the state shown in Fig. 7*c*, in the absence of the coupling spring. Uncopying can therefore also be done with arbitrarily little energy dissipation. The time-modulated potential scheme described above has implications for certain biological systems^{20,21}.

We can think of the two wells in Fig. 7 as part of a much longer chain of wells, in which the right-hand well passes its information state on to another well further to the right. In that case we have a communication link and can move a bit through a chain of given length with as little dissipation as desired.

We have explained that reversible computation can be carried out with very little dissipation only if done slowly. How slow is 'slowly'? Such a question can readily be answered for a particular implementation of a reversible computer, for example Likharev's Josephson junction scheme^{15,16}, with specified circuit parameters. There is no obvious way to provide a more universally applicable answer, because there are no clear constraints on the timescale of computational kinetics. After all, in principle we could use the internal degrees of freedom of the nucleus, or the W and Z particles, in computational devices.

Reversible computers give more outputs than are needed at many of the logic stages, and as a result they are unreasonably demanding in terms of hardware. To an appreciable extent, the

need for hardware can be traded for extra steps (C. H. Bennett, preprint). I feel that the need for extra hardware or extra steps, combined with the fact that we are normally not particularly interested in carrying out each step slowly, makes reversible low-dissipation computers largely a matter of purely conceptual interest. (This opinion is not shared by some others in the field.) Energy consumption in real computers is important, of course, because it determines the density with which we can pack circuits, which in turn controls the delay encountered by signals in going from one logic stage to another. But, as indicated in Fig. 1, a great many engineering innovations are required before we reach the kT limit at which reversible computation becomes significant.

Measurement

We will not discuss measurement in detail, in view of recent definitive discussions by Bennett^{7,18,19} and others (refs 22-24 and S. Lloyd, preprint). Furthermore, the word measurement is ambiguous, and remains so even in some sophisticated discussions of the concept^{25,26}. Some believe that involvement of a human observer is an essential ingredient. I believe that James Watt's invention of the centrifugal governor showed that the human mind did not have to participate. The thermostat in your home makes a measurement, even when you are not there. Other authors, although not arguing for human involvement, do invoke the necessity of recording or registry in a macroscopic piece of apparatus, which I interpret as at least more reasonable. Nevertheless, at a time when we can manipulate and observe individual atoms and when we know how, in principle, to use apparatus modelled on the genetic code to do computation^{7,27}, it is not clear that macroscopic apparatus is essential.

Figure 8 typifies the measurement process, in which a system to be measured, *S*, is brought into interaction with a meter *M*. As a result of this interaction, *M* is changed to a new state *M'*, which provides information about *S*. The measured system may (but does not necessarily have to) change to a new state *S'*. The meter *M* and the system *S* (or *S'*) are typically then decoupled. This decoupling is not essential; for example, the home thermostat is a heavily damped system which reflects only the recent thermal history of its environment. But the decoupling is typical of many measurement operations, particularly of those found in theoretical discussions. It is also characteristic of the discussions of measurement related to Maxwell's demon.

A coupled time evolution, as shown in Fig. 8, does not require dissipation. Having gone through such a time evolution it is possible in principle, to reverse the process and undo the measurement process that has just occurred. (In fact, it is not really clear, particularly in a quantum mechanical system, that reversal of motion is a physically executable operation. Here, we shall not confront this possibility, but instead we make the prevalent and unwarranted assumption that it presents no problem.) But undoing the measurement makes the measurement useless, which is not desirable. Usually the measurement prompts a subsequent action or response; for example, in the original version of Maxwell's demon⁴ a trap door is opened or

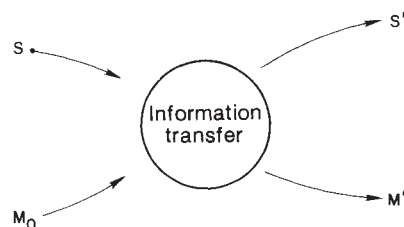


Fig. 8 Schematic representation of the measurement process. A system *S* is coupled to a meter, which is in a standardized initial state *M*₀. After a coupling is established and then removed, the meter is left in a state *M'*, providing information about *S*. The final state of the measured system may change to a new state *S'*, but such a change is not essential.

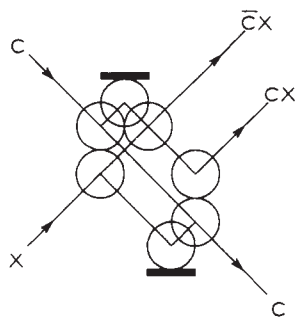


Fig. 9 Two inputs C and X interact by the direct collision of billiard balls, and their subsequent reflection by mirrors. The incident balls take an undeflected course if only one is present, but alternative trajectories if a collision occurs. The output labelled CX is present if and only if both C and X are present. $\bar{C}X$ is present if and only if C is absent and X is present.

closed, depending on the velocity and direction of an approaching molecule. In Szilard's²⁸ later version of the demon, the position of a molecule, that is, whether it is to the left or right of a barrier in the middle of a vessel, is used to control a piston against which the molecule does work.

Once the state of the meter in Fig. 8 has been used, we then need to reset it to a standardized state, for its next (or first) use. This involves erasure of information and, as already pointed out, requires a minimum energy dissipation. It is this dissipation associated with resetting which saves the second law in the operation of Maxwell's demon; we need not look for the source elsewhere^{7,18,19}. Unfortunately, Brillouin³, Gabor²⁹, and others who followed them, were not aware of this. To save the second law they looked for dissipation in the information transfer step and found dissipative ways of performing this step. It is remarkable that neither these authors nor the many follow-up discussions asked whether this was a minimally dissipative procedure. In fact, the arguments invoking dissipative information transfer were so well accepted that the trail leading to them became a topic in the history of science³⁰.

Frictionless and quantum computers

The reversible computers discussed up to now have friction and noise. These properties may be inevitable in the real world, but are they essential to the computational process? A number of authors have analysed both classical and quantum frictionless systems that supposedly carry out computation. These include Benioff, Deutsch, Feynman, Fredkin, Likharev, Margolus, Peres, Toffoli and Zurek; see refs 18 and 31 for citation details. Figure 9 illustrates one such scheme, in which collisions between classical billiard balls are used to execute logic¹⁰. Information is represented by the absence or presence of billiard balls, and the interaction of signal streams is obtained by collisions. Reflecting walls, or 'mirrors', are also needed to guide balls towards further collisions.

This is a system that requires perfection. A slightly misdirected ball will go increasingly further off course with time until eventually the systems ceases to work altogether. In fact, this is a chaotic system; collisions between balls will cause an exponential, rather than linear, growth in the error. Thus, any source of error—incorrect initial launching velocity or position, a misaligned mirror, friction or noise—will cause the system to fail. This is therefore a pathological form of computation. Ordinary computers continually re-standardize signals towards their ideal specification, for example pushing electrical signals toward B+ or ground ('1' or '0'). Such re-standardization, however, discards information, specifically information about the error. As we have seen, to discard information requires dissipation; the phase space of the information-bearing degree of freedom is compressed in the error correction process, which has to be compensated by a phase-space expansion, or entropy increase, in

the irrelevant degrees of freedom. In other words, these are heated. Dissipation, therefore, is essential to a real computer, which cannot count on perfection; it is not just an unavoidable nuisance.

Deutsch has proposed a computer³² that can be started in a coherent quantum mechanical superposition of different initial states. Judicious measurements at the end can yield simultaneous information about results corresponding to the superposed initial states, and thus constitute a form of parallelism analogous to the observation of a diffraction pattern in a two-slit experiment for a quantum-mechanical particle. The diffraction pattern depends, in detail, on the length of, and the potential encountered along, both paths. Through an inelastic event, however, in which the particle gives up energy to other degrees of freedom along one (or both) of the two paths, the interference pattern is destroyed and the particle $\bar{c}$ can be viewed as having come along only one of the two possible paths. Thus, a computation that is to yield a result that depends on two (or more) simultaneous computational trajectories requires a completely phase-coherent quantum mechanical evolution, in the absence of friction. But, as indicated, that in turn requires unrealizable perfection for the parts of the computer.

The existing descriptions of quantum mechanical computation in the literature have two additional deficiencies, albeit of lesser significance because there is some chance that they can be addressed. First, consider Fig. 4a and its analogy with an electron moving in a periodic lattice. In quantum mechanics, if a one-dimensional lattice is not perfectly periodic, the electrons cannot propagate freely, but will be localized. The transmission probability will decrease exponentially as the length of the chain is increased. That this is not a totally irremediable defect is suggested by the quantized Hall effect³³, in which the states propagating along the edge of a sample can be transmitted without reflection by irregularities. The second problem with the quantum mechanical computers comes from the lack of explicit description of the initial program-loading process. But I would suggest that is more likely to represent a deficiency in the published descriptions, rather than constituting an intrinsic difficulty.

Communications channel

There is a widespread belief that it requires an energy expenditure of at least $kT \ln(2)$ to transmit a bit through a channel. This idea originates from Shannon, who states² (italics added here for emphasis): 'An important *special case* occurs when the noise is *added* to the signal and is independent of it (in the probability sense)'. In that case, Shannon finds

$$C = W \ln [N^{-1}(P + N)] \quad (1)$$

Here C is the channel capacity, W the bandwidth, P the average received power and N the average noise power. Thermal noise, for a classical transmission line with additive equilibrium noise, is given by $N = kTW$. Equation (1) then yields a minimum for P/C , at small P , given by $(P/C) = kT \ln(2)$. Thus an energy of $kT \ln(2)$ per bit is required, but it is not clear that this has to be dissipated. Equation (1) has an obvious credibility. N determines the separation between distinguishable signals; the larger N , the fewer the number of distinguishable messages. But communication does not need wave propagation: mail is communication; physical transport of a PC disk is communication. Communication, furthermore, does not have to use degrees of freedom in which the effects of noise are simply added to that of the signal in a linear transmission process. There is an obvious advantage in using states of local stability to denote information (as done in our discussion of potential wells), in contrast to the use of states in which noise causes unhindered diffusion from one possible state to another. Furthermore, even if channel capacity expressions describe the energy required in the message, does the energy have to be dissipated? As shown above, Shannon understood at least some of these limitations. Later

authors were less perceptive, with two exceptions: the use of matter, rather than waves, to transmit information has been discussed by Pendry³⁴ and Bekenstein³⁵. A scheme in which the transfer of matter is used to transmit with arbitrarily little dissipation has been described by this author^{24,36}.

As suggested above, we can utilize nonlinear wave propagation systems. Kinks in a sine-Gordon soliton chain can be transmitted with as little dissipation as desired²⁴. Most clearly and simply, however, a chain built out of the time-modulated wells of Fig. 7 will accomplish the desired goal.

I should emphasize that much of Shannon's pioneering work² is beyond criticism. I do not question his general formulation in which channel capacity is expressed in terms of the statistics of message choice at the transmitting end, and in terms of the probabilities with which a transmitted symbol turns up as a received symbol. My reservations arise only from the limited applicability of a particular physical model for the calculation of error probabilities incurred in transmission.

Conclusions and extensions

Limits (of the order of kT) in both computation and communication, which appeared intuitively obvious and reasonable, are in

fact avoidable. Similarly, it has been suggested in a number of sources that the quantum mechanical uncertainty principle provides an estimate for minimal energy dissipation in the case of fast switching events. The quantum mechanical reversible computers to which we have alluded present counter-examples to that suggestion³⁷. The real limits are presumably related to cosmological questions. For example, how many degrees of freedom in the Universe can we bring together effectively to make a computer or memory? Or, as another example, consider the typical and inevitable deterioration processes which beset real machinery: oxidation, diffusion, evaporation, electromigration, irradiation by cosmic rays and α -particles. Can their effects on computer reliability be offset to an arbitrary degree? Such questions raise serious doubts about the implementability of continuum mathematics in our real physical Universe. The corresponding implications for the nature of the laws of physics, which ordinarily are expressed in classical mathematical form, have been discussed elsewhere^{38,39}. Such questions are also, in my speculative opinion, related to the origin of irreversibility²⁴. These subjects, however, are currently not developed beyond casual speculation and are unsuitable for this brief review.

1. Keyes, R. W. *IBM J. Res. Dev.* **32**, 24–28 (1988).
2. Shannon, C. E. *Bell Syst. tech. J.* **27**, 379–423, 623–656 (1948).
3. Brillouin, L. *Science and Information Theory* 267–286 (Academic Press, New York, 1956).
4. Maxwell, J. C. *Theory of Heat* 1st edn (Longmans Green, London, 1871).
5. Landauer, R. *IBM J. Res. Dev.* **5**, 183–191 (1961).
6. Bennett, C. H. *IBM J. Res. Dev.* **17**, 525–532 (1973).
7. Bennett, C. H. *Int. J. theor. Phys.* **21**, 905–940 (1982).
8. Landauer, R. in *Der Informationsbegriff in Technik und Wissenschaft* (eds Folberth, O. G. & Hackl, C.) 139–158 (Oldenbourg, Munich, 1986).
9. Wright, R. *The Atlantic* **261**, 29–44 (April, 1988).
10. Fredkin, E. & Toffoli, T. *Int. J. theor. Phys.* **21**, 219–253 (1982).
11. Landauer, R. *Ber Bunsenges* **80**, 1048–1059 (1976).
12. Landauer, R. *Int. J. theor. Phys.* **21**, 283–297 (1982).
13. von Neumann, J. *Non-linear Capacitance or Inductance Switching, Amplifying and Memory Organs* (US Patent 2,815,488).
14. Goto, E. *J. electr. Commun. Engrs Japan* **38**, 770–775 (1955).
15. Likharev, K. K. *Int. J. theor. Phys.* **21**, 311–326 (1982).
16. Likharev, K. K., Rylov, S. V. & Semenov, V. K. *IEEE Trans. Magn.* **21**, 947–950 (1985).
17. Landauer, R. & Büttiker, M. *Phys. Scripta* **T9**, 155–164 (1985).
18. Bennett, C. H. *IBM J. Res. Dev.* **32**, 16–23 (1988).
19. Bennett, C. H. *Sci. Am.* **255**, 107–116 (1987).
20. Kuhn, H. *IBM J. Res. Dev.* **32**, 37–46 (1988).
21. Kondepudi, D. K., Moss, F. & McClintock, P. V. E. *Physica D* **21**, 296–306 (1986).
22. Zurek, W. H. in *Frontiers of Nonequilibrium Statistical Physics* (eds Moore, G. T. & Scully, M. O.) 151–161 (Plenum, New York, 1986).
23. Lubkin, E. *Int. J. theor. Phys.* **26**, 523–535 (1987).
24. Landauer, R. in *Signal Processing* (ed. Haykin, S.) (Prentice-Hall, New Jersey, in the press).
25. Greenberger, D. M. (ed. *New Techniques and Ideas in Quantum Measurement Theory* (Ann. New York Acad. Sci., 1986).
26. Roth, M. & Inomata, A. (eds) *Fundamental Questions in Quantum Mechanics* (Gordon & Breach, New York, 1986).
27. Bennett, C. H. & Landauer, R. *Sci. Am.* **253**(7), 48–56 (1985).
28. Szilard, L. *Z. Phys.* **53**, 840–856 (1929).
29. Gabor, D. in *Progress in Optics* Vol. 1 (ed. Wolf, E.) 109–153 (North-Holland, Amsterdam, 1961).
30. Daub, E. E. *Hist. Phil. Sci.* **1**, 213–227 (1970).
31. Landauer, R. *Found. Phys.* **16**, 551–564 (1986).
32. Deutsch, D. *Proc. R. Soc. Lond. A* **400**, 97–117 (1985).
33. Streda, P., Kucera, J. & MacDonald, A. H. *Phys. Rev. Lett.* **59**, 1973–1975 (1987).
34. Pendry, J. B. *J. Phys. A* **16**, 2161–2171 (1983).
35. Bekenstein, J. D. *Phys. Rev. A* **37**, 3437–3449 (1988).
36. Landauer, R. *Appl. Phys. Lett.* **51**, 2056–2058 (1987).
37. Sturge, M. D. in *No Way* (eds Davis, P. J. & Park, D.) 111–138 (Freeman, New York, 1987).
38. Landauer, R. *Speculations in Science and Technology* **10**, 292–302 (1987).
39. Landauer, R. *Phys. Scripta* **35**, 88–95 (1987).

ARTICLES

Mantle-derived fluids in diamond micro-inclusions

O. Navon, I. D. Hutcheon, G. R. Rossman & G. J. Wasserburg

Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA

Micro-inclusions in diamonds from Zaire and Botswana differ in composition from the more common large inclusions of the peridotitic or eclogitic assemblages. These sub-micrometre inclusions resemble potassic magmas in their composition, but are enriched in H_2O , CO_3^{2-} and K_2O and depleted in MgO . This composition represents a volatile-rich fluid or melt from the upper mantle, which was trapped in the diamonds as they grew.

The transport of volatiles and incompatible elements in the upper mantle is dominated by the migration of fluids (melts or volatile-rich fluids). Although well studied processes such as diamond genesis and mantle metasomatism have been associated with the presence of such fluids^{1–3}, direct evidence on deep mantle fluids is scarce and their exact nature is still debated^{3,4}. Because of its mechanical strength and chemical inertness, diamond is probably the best material for transporting such fluids to the surface. Nevertheless, no conclusive, direct observation of fluid inclusions in diamonds has been reported⁵. Infrared (IR) spectroscopic studies have indicated the presence of water and carbonate in sub-micrometre inclusions in cubic diamonds⁶

and in overgrowths (coats) on the clear, inclusion-free cores of coated diamonds⁷. Instrumental neutron activation analysis (INAA) of a coated diamond indicated patterns enriched in light rare-earth elements and led Bibby⁸ to suggest that the inclusions contain melt. That study was unable, however, to determine the major-element composition of the included matter.

We have studied the composition and mineralogy of the micro-inclusions in cubic and coated diamonds from Zaire and Botswana, where such diamonds are commonly found. These inclusions are rich in H_2O , CO_3^{2-} , SiO_2 , K_2O , CaO and FeO . Their bulk composition resembles that of potassic magmas, such

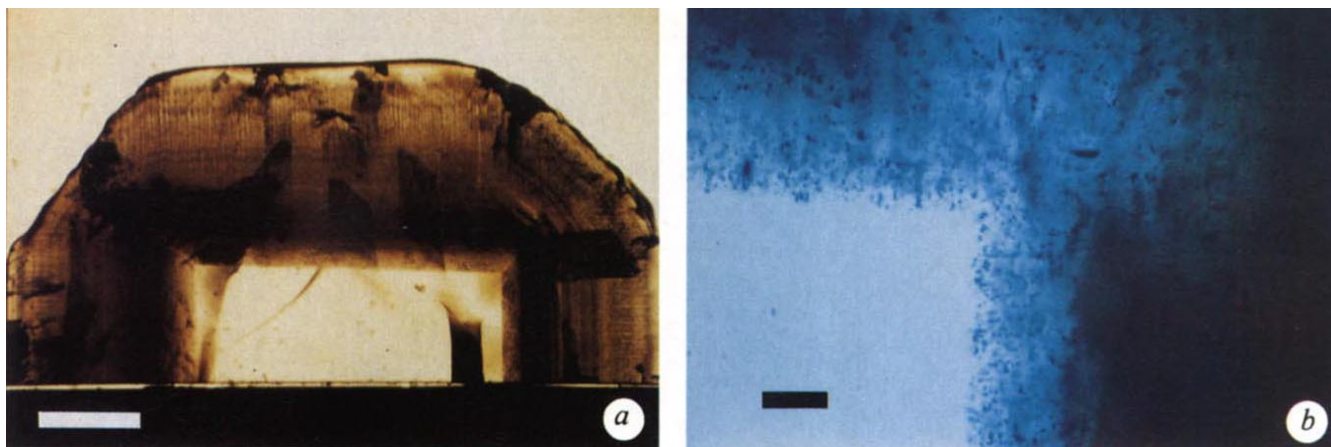


Fig. 1 Micro-inclusions in coated diamonds. *a*, A polished section of CTP LB, a coated diamond from Zaire. The transparent core is an octahedral diamond; the cuboid habit of the coat can be recognized by the development of cuboid faces at the corners. The dark parallelogram forms that run through the coat and the core are cracks. They do not disturb the concentric banding of inclusion-rich and inclusion-poor zones and are clearly late. Transmitted light, scale bar indicates 1 mm. *b*, High magnification of the core-coat boundary. Individual inclusions are of sub-micrometre size. The few larger, dark features are due to chipping of the diamond surface during polishing. Transmitted light, blue filter, scale bar represents 10 μm .

as kimberlites and lamproites, but is depleted in MgO, extremely enriched in K_2O and has a high $\text{K}_2\text{O}/\text{Al}_2\text{O}_3$ ratio. IR spectra indicate a mineralogical assemblage that includes hydrated sheet silicates, carbonates and phosphates; in most cases, some molecular CO_2 is also present. We infer that the bulk composition of the inclusions represents a volatile-rich fluid or melt trapped by the diamonds during their growth.

Samples and technique

The six cubic and nine coated diamonds studied here (Table 1) are translucent to opaque and vary in colour from greyish-brown to green to yellowish-green; many exhibit trigonal etch pits (trigons) and channels on their faces. The cubic diamonds fit the description of variety III and the coated diamonds are of variety IV according to the classification of Orlov⁹. X-ray topography^{10,11} has shown that coats of coated diamonds and many cubic diamonds possess a fibrous internal structure in the (111) direction. The cores of the coated diamonds are transparent single crystals of octahedral shape.

Optical examination of the coats and the cubic diamonds reveals numerous sub-micrometre, sub-rounded inclusions, completely enclosed within optically continuous diamond (Fig. 1). In many cases the inclusions are arranged in nearly parallel arrays of isolated inclusions, perpendicular to the growth faces of the diamond. The arrays are separated by 20–80- μm thick, inclusion-free zones. In other cases the inclusions are distributed randomly. The outer rims of the diamonds are usually free of inclusions¹², while the interiors show alternating concentric bands resulting from variations in the number density of inclusions.

The diamonds were cut and polished into ~0.4–1-mm thick wafers for spectroscopic and chemical analysis. Infrared absorption spectra were obtained by Fourier transform infrared spectroscopy (FTIR); after infrared characterization the wafers were mounted in epoxy in brass holders and elemental contents were determined by electron probe microanalysis (EPMA), secondary ion mass spectrometry (SIMS) and analytical scanning electron microscopy (SEM). Fragments of some samples were analysed using INAA by the generous cooperation of B. Spettel and H. Palme at the Max-Planck-Institut für Chemie. All samples were cleaned with hot concentrated HF and HCl and were rinsed in water and ethanol to remove surface contamination. Infrared (IR) spectra were collected on a Nicolet 60SX FTIR¹³ using 200–600- μm apertures. EPMA data were collected on a JEOL

733 EPMA with a 50-nA, 15-kV beam, defocused to 20 μm . Four adjacent points, covering a 40 $\times$ 40- μm area, were analysed with total counting times of 800–1,200 s. Each of the analysed areas typically contains a few hundred inclusions; an EPMA analysis therefore represents an average composition. SEM analyses of individual inclusions were made with a JEOL JSM 35CF equipped with a Tracor Northern TN-5500 energy dispersive spectrometer (EDS). A 0.075-nA, 15-kV beam was rastered over a 0.3 $\times$ 0.3- μm area centred over an inclusion. Electron probe and SEM data were reduced using a ZAF procedure¹⁴. The accuracy of the correction procedure was verified by analysing a graphite pellet containing 2% of glass powder of known composition.

SIMS data were obtained with PANURGE, a modified Cameca IMS-3F¹⁵. A 10-nA, 14.5-kV O^- beam was focused to a ~40- μm diameter spot. Positive secondary ions were analysed using a mass resolving power of $m/\Delta m = 2,000$, sufficient to resolve all significant molecular interferences. Detection limits were <5 parts per 10^9 for Na, Mg, K and Ca, and <500 parts per 10^9 for Al, Si, Ti and Fe; precision was much higher than that achieved by EPMA. A SIMS analysis also represents an average of a few hundred inclusions.

A key experimental issue was the determination of the SIMS ion yields, in order to calculate the concentrations of the different elements in the sputtered volumes from measured secondary ion intensities. Because diamonds containing known concentrations of the elements of interest were unavailable, ion intensities were calibrated against EPMA concentrations. Ion intensities were first normalized to the intensity of $^{12}\text{C}^+$ to correct for variations in ion yields associated with changes in operating conditions. A working curve was constructed for each element by correlating the normalized secondary ion intensities with concentrations determined by EPMA on the same inclusion-rich area of a diamond. The accuracy of the working curves was evaluated using EPMA and SIMS analyses of glass-graphite pellets containing 0.4 and 2% of glass powder of known composition. Ion yields are listed in Table 1, together with the associated 2σ errors. Errors are much larger than expected on the basis of counting statistics alone and arise chiefly from the heterogeneous distribution of inclusions in the diamonds coupled with the small difference between the analytical volumes of the EPMA and SIMS.

Each of the techniques we used samples a different volume element. The X-ray volume excitation for SEM analysis is

Table 1 Metal oxides, water and carbonate in micro-inclusion-bearing diamonds

Sample	CTP 6268 Oct.†	CTP L0 Oct.	CTP L6 Oct.	CTP LB Oct.	CTP Z4 Oct.	CTP MM1 Oct.	GRR 1503 Oct.	GRR 1504 Oct.	GRR 1508 Oct.	GRR 861.2 Cube	GRR 1155 Cube	GRR 1515 Cube	GRR 1517 Cube	GRR 1518 Cube	GRR 1519 Cube	Relative ion yields*
Source	Jwaneng, Botswana	Zaire	Zaire	Zaire	Zaire	Mbuji-Mayi Zaire	Unc.‡	Unc.	Unc.	Unk.‡	Unk.	Unc.	Unc.	Unc.	Unc.	
Points§	4	5	2	2	3	2	7	8	10	5	10	7	3	2	2	
SiO ₂	31.9	41.2	43.3	34.6	67.7	40.4	35.6	42.3	42.4	51.1	53.6	45.1	30.3	42.4	45.9	800±150
TiO ₂	4.2	2.4	2.5	2.1	2.0	2.9	2.8	2.6	2.7	2.4	4.0	2.3	3.4	2.9	2.6	2,500±1,600
Al ₂ O ₃	2.9	6.1	5.4	5.6	5.9	4.5	3.3	4.9	4.9	5.4	4.3	4.6	5.3	4.4	4.8	11,400±2,800
FeO	15.7	5.0	5.6	4.9	3.3	7.2	8.3	11.1	6.1	6.8	6.6	10.1	5.0	8.0	8.8	800±400
MgO	5.7	2.8	3.8	2.3	1.3	4.6	6.1	4.6	3.6	5.7	2.6	8.0	4.3	4.9	4.9	4,300±1,100
CaO	10.5	10.7	10.6	12.3	1.6	13.9	16.8	7.8	11.9	8.6	7.2	7.6	18.7	9.8	12.4	10,200±1,600
Na ₂ O	2.6	3.0	2.9	3.5	1.0	3.8	2.9	2.3	2.4	2.1	1.0	4.8	2.7	3.4	3.4	57,000±20,000
K ₂ O	21.4	23.7	20.8	29.7	12.3	17.7	18.6	19.4	21.1	11.6	15.5	12.4	25.2	19.3	12.1	35,000±4,300
Total metal oxide (p.p.m.)¶	1,195	433	211	247	1,412	107	559	1,207	508	80	551	628	22	118	99	
H ₂ O (p.p.m.)#	407	191	118	140	294	165	282	619	269	98	168	241			148	
CO ₂ (p.p.m.)	600	79	44	66	107	89	292	135	133	67	173	139			44	
H ₂ O/(H ₂ O+CO ₂) (molar ratio)	0.5	0.8	0.8	0.8	0.7	0.6	0.9	0.8	0.7	0.6	0.7			0.8		

* SIMS ion yields are in (counts per second/p.p.m.)_{element}/(c.p.s./p.p.m.)_{carbon}. Errors are estimated as 2σ of the slope of the EPMA-SIMS correlation lines.

† Oct. = coated octahedron.

‡ Unc. = uncertain source, probably Zaire, according to diamond morphology (J. W. Harris, personal communication); Unk. = unknown.

§ Number of SIMS analyses on each diamond.

|| Average major oxide composition (wt %) of the volatile-free fraction of the inclusions, calculated by normalizing the total the eight metal oxides to 95%; on average, P₂O₅, BaO, SrO and the rare earth elements account for ~5% of the metal oxide fraction. Standard deviation of individual analyses from the average diamond composition is typically less than 10%.

¶ Total concentration of the above eight oxides determined by SIMS (in p.p.m.). Totals represent averages of all points analysed on each diamond. Standard deviations are typically 30%, resulting from variations in the abundance of inclusions; however it is possible that in few cases the surface analysis is not representative of the amount of matter in the whole diamond.

Concentration of water was calculated using $\epsilon_{3,420} = 80$ l per mol-cm (liquid water, ref. 37). Concentration of carbonate, given in terms of p.p.m. CO₂, was estimated using $\epsilon_{1,430} = 250$ l per mol-cm (calcite, G. Fine, personal communication). These values are believed to be representative of the absorption by the hydrous and carbonate species in the inclusions. Absorption coefficients of other carbonates (for example, Na₂CO₃, gaylucite) differ by less than 30%. The water absorption is also believed to be accurate to ~30%. The precision of the IR spectral data is ±25 p.p.m. H₂O and ±40 p.p.m. CO₂.

~2 µm in diameter, and total oxide concentration is typically a few per cent. SIMS or EPMA analyses sample the surface (1 µm depth) of a ~40-µm diameter area and total oxide concentration is a few hundred p.p.m. IR spectroscopy samples the entire wafer thickness and INAA yields element concentrations for the bulk sample. Assuming that all oxides are in the inclusions, the data are interpreted to obtain the bulk concentration of a species in a given volume (including the diamond) and the average composition of the inclusions in that volume or, in the case of SEM analysis, the composition of individual inclusions.

The constancy of both the elemental ratios from SIMS analysis and the CO₂²⁻/H₂O ratios from FTIR analyses of individual diamonds indicate that, although the amount of trapped material is variable, the average composition of the inclusions in different zones of any one diamond is reasonably uniform. Thus, it is meaningful to compare compositional data measured by different techniques on different portions of a single diamond, as long as at least one element is measured by each technique. SIMS, SEM and INAA data can be normalized against the concentrations of K or Na that were determined by all three techniques; the comparison of SIMS and FTIR data is problematic, as we lack a common component.

Composition and mineralogy

The SIMS analyses of 15 diamonds are summarized in Table 1. The average composition of the volatile-free fraction of the micro-inclusions in each diamond is presented as the oxide weight per cent of the eight major constituents, normalized to 95%; the remaining 5% consists of other minor oxides measured only by EPMA, SEM and INAA. The total concentration of the eight metal oxides in each diamond is also shown.

Several features are apparent in the data. The total concentration of metal oxides is widely variable (20–1,270 p.p.m.), which we attribute to variation in the number density of inclusions.

In contrast, the average composition of the micro-inclusions in the various diamonds is broadly similar (Fig. 2). All inclusions are rich in SiO₂, K₂O, CaO and FeO. Compositions of the metal oxide fraction in most diamonds vary within the following ranges: SiO₂, 30–53 wt %; K₂O, 12–30%; CaO, 8–19%; FeO, 6–11% (total iron as FeO); Al₂O₃, 3–6%; MgO, 2–6%; TiO₂, 2–4%; and Na₂O, 1–5%. Not shown in the table but also present are P₂O₅, 1–4%; BaO, 1–4%; SrO, 0.7–1.5%; La₂O₃, 0.1–0.3%; Ce₂O₃, 0.3–0.5%; and smaller amounts of other rare-earth elements, Mn, Th and U. K₂O/Al₂O₃ ratios are very high (2–7) and chondritic normalized Ce/Eu ratios vary between 10–21. SEM analyses of individual inclusions found 1–3% chlorine; no sulphur was detected and its content relative to the total oxide content is less than 1%.

Inclusions in three diamonds have compositions clearly distinct from that of the general population. Inclusions in CTP Z4 are much lower in CaO, MgO, K₂O and Na₂O and higher in their SiO₂ content (see Fig. 2). Inclusions in CTP 6268 (the only sample from Jwaneng, Botswana) contain 1.5 to 2 times more FeO than inclusions in other diamonds and also have high TiO₂ and MgO content. Inclusions in GRR 1515 are rich in MgO. These three samples also show distinct features in their IR spectra.

SEM analysis of individual inclusions revealed no monomineralic inclusions, but recorded the same Si, K, Ca, Fe-rich composition found by SIMS. The average composition of the inclusions in a single diamond agrees with the SIMS results.

IR spectra of inclusion-bearing zones (Fig. 3a) exhibit absorption bands characteristic of: (1) pure diamond; (2) type IaA nitrogen¹⁶ (1,280, 1,220 and 480 cm⁻¹, corresponding to 500–1,100 p.p.m. nitrogen in the diamond lattice); and (3) impurities such as carbonate and water⁷. Small peaks at 1,365 (N in Fig. 3) and 3,107 cm⁻¹ (C–H) are due to nitrogen platelets¹⁶ and hydrogen¹⁷ in the diamond matrix, respectively. The absorption features produced by the micro-inclusions were obtained as the

difference between the spectra of the inclusion-bearing zones and a spectrum of clear diamond with the same amount of nitrogen. The difference spectra of the inclusion-bearing zones in all diamonds are similar and exhibit the same bands.

The intensities of the two characteristic carbonate bands at 1,430 and 876 cm^{-1} (CO_3^{2-} in Fig. 3) in different diamonds are strongly correlated. The narrow band at 876 cm^{-1} is in the same position ($\pm 2 \text{ cm}^{-1}$) in all diamonds; this position is characteristic of calcite and is distinct from that in dolomite or magnesite. The two bands at 575 and 605 cm^{-1} (P) are due to phosphate absorption and are similar in position and shape to the doublet of apatite. Apatite has also been found in a TEM study of micro-inclusions in a coated diamond¹⁸. Molecular CO_2 is present in most but not all diamonds and gives rise to the band at 2,350 cm^{-1} . The strong bands at 1,630 and 3,420 cm^{-1} are due to H-O-H bending and O-H stretching of water, respectively. Sharp narrow bands, typical of hydroxyl groups, are not found around 3,600 cm^{-1} and must be weaker than the 3,420 cm^{-1} band, typical of the stretching vibration of molecular water. In conjunction with the relative height of the H-O-H band and the very good correlation between the intensities of the 1,630- and 3,420- cm^{-1} bands, this suggests that most water is present as H_2O molecules. No frequency shift towards ice absorption was detected at liquid nitrogen temperatures, suggesting that water is not present as a bulk aqueous phase.

The bands at 475, 525, 685, 1,000 and 1,100 cm^{-1} (S) are characteristic of silicate absorption. Their intensities are inter-correlated and are also correlated with the intensity of the water bands and of the band at 840 cm^{-1} (S1). The two bands at 785 and 812 cm^{-1} (S2) show a strong inter-correlation and a weaker correlation with the silicate bands. The S1 and S2 bands are in the spectral range characteristic of metal-OH-metal bands of micas and clay minerals. The position of all the silicate bands (S, S1, S2) and the good correlation between the intensities of the main silicate and water bands are best explained by the presence of heavily hydrated clay minerals. The presence of more than one silicate phase is indicated by the lack of correlation between the intensities of the S1 and S2 bands. The inclusions in diamonds CTP Z4 and CTP 6268 are of extreme chemical compositions (Fig. 2) and show contrasting and extreme IR spectra. CTP Z4 has intense S2 and 1,100- cm^{-1} bands. CTP 6268 has no S2 bands, but has intense S1 and 1,000- cm^{-1} bands. The intensities of the different S bands in these two diamonds do not exhibit the good correlations observed for other diamonds. The intensities of the carbonate

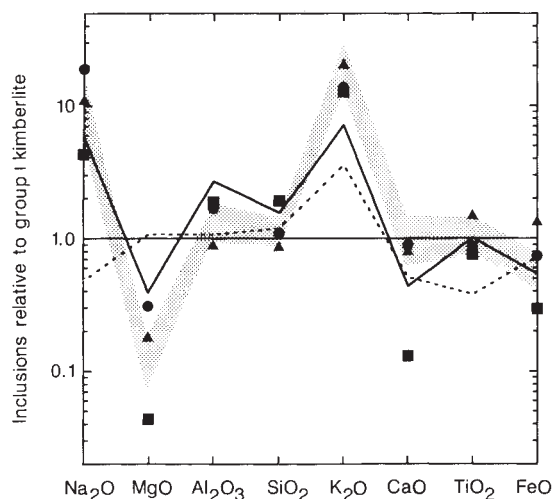


Fig. 2 Major-element abundances in diamond micro-inclusions relative to average group I kimberlite³⁸. Stippled area represents all diamonds of Table 2 except CTP Z4 (■), CTP 6268 (▲) and GRR 1515 (●). Solid line indicates an average for lamproite³⁹; dashed line represents group II kimberlites³⁸.

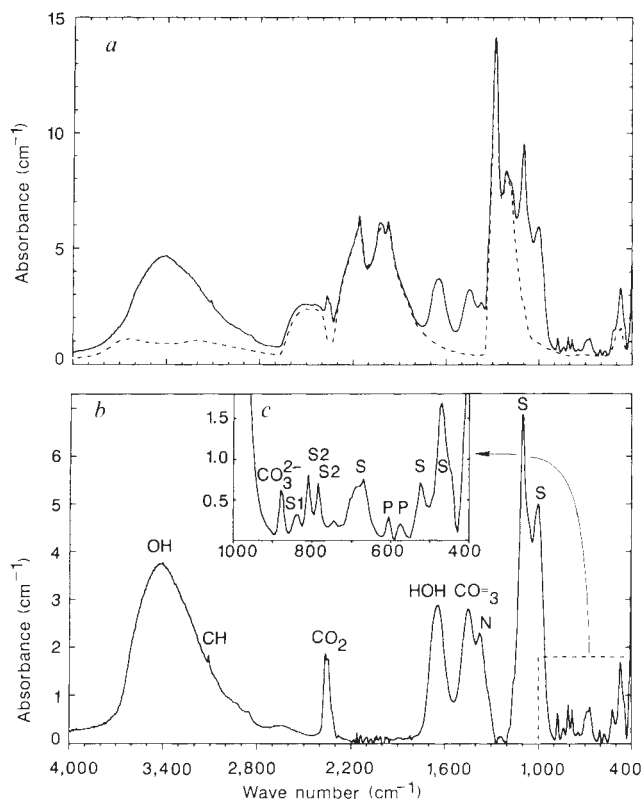


Fig. 3 *a*, IR spectrum of turbid diamond GRR 1515 (solid line) and a synthetic, composite, reference spectrum of a pure diamond plus the appropriate amount of type IaA nitrogen centres (dashed line). *b*, IR absorbance resulting from micro-inclusions in GRR 1515, obtained as the difference between the two spectra in *a*. *c*, Details of the low wavenumber region of *b*.

bands, the silicate-water bands and the phosphate bands show only a weak correlation, probably because of varying proportions of the three phases in different diamonds. Olivine, pyroxene and garnet have some characteristic lines which were not detected and therefore cannot be significant components of the inclusions.

The total water and carbonate concentrations for each diamond studied by IR spectroscopy are presented in Table 1; concentrations of water and carbonate were calculated using absorption coefficients of pure water and calcite (see footnotes to Table 1). The concentrations of both species are quite variable, spanning the range ~44–600 p.p.m., with water being more abundant in all diamonds (molar proportions). Water-to-carbonate ratios vary among the different samples but are uniform throughout individual diamonds. As SIMS and FTIR sample different components in different volumes, an accurate determination of the relative proportions of water or carbonate in the inclusions is difficult. Accurate estimation will be achieved only when the same volume is analysed by FTIR and a bulk chemical analysis (such as INAA). At present, we can only estimate that water and the CO_2 component of the carbonate account together for ~40% of the total weight of the inclusions in most diamonds (Table 1). The absorption coefficient of molecular CO_2 depends strongly on pressure and its concentration cannot be well constrained. Rough estimates suggest that the CO_2 content is much less than the CO_3^{2-} content.

Mantle-derived fluids

Micro-inclusions in cubic and coated diamonds are rich in H_2O , CO_3^{2-} , SiO_2 , K_2O , CaO and iron oxide. Their composition and IR spectra are not compatible with any single mineral phase or with any combination of the common phases present in large diamond inclusions¹⁹ (such as olivine, garnet and pyroxene).

The bulk composition of the micro-inclusions does resemble that of potassic magmas, especially those of lamproites and group II kimberlites, but some important distinctions exist. Our estimate of the amounts of water and CO₂ (in carbonates) is almost an order of magnitude higher than the volatile content of lamproites and kimberlites. MgO is depleted and the enrichment of potassium and other incompatible elements is even higher than in lamproites or kimberlites.

The question arises as to whether the bulk composition of the inclusions is representative of the growth environment of the diamonds. The concentric zoning patterns are interpreted as growth features^{10,11} and suggest that the micro-inclusions were formed during the growth of the diamonds. It is not possible to prove whether the chemical composition of the inclusions is primary or reflects secondary alteration. However, the inclusions appear to be completely enclosed within an optically continuous diamond matrix and kimberlitic penetration into the diamond coat is limited to less than a few micrometres¹⁴. In addition, the composition of the inclusions reported here is different from either the kimberlite (Fig. 2) or kimberlitic alteration products (for example, magnesian clays²⁰). Thus we believe that the inclusions have retained their original bulk composition. The current mineral assemblage of hydrated sheet silicates, carbonates and phosphates may be the result of low-temperature, closed-system crystallization of the original trapped material.

The high content of volatiles and incompatible elements and the uniform average composition of the micro-inclusions suggest that the original material trapped by the diamond was a fluid, either a volatile-rich fluid or a melt. Can such fluids, containing ~40% volatiles, exist in the upper mantle? Ryabchikov and Boettcher²¹ have shown that the amount of solute in potassium-rich hydrous fluids in equilibrium with phlogopite + forsterite increases with increasing pressure and reaches 50% at 30 kbar and 1,000 °C. Ellis and Wyllie²² have estimated that, at 50 kbar pressure, near-solidus liquids in equilibrium with CO₂-H₂O fluid, forsterite and enstatite may contain 30% CO₂ and 15% H₂O. Based on these data and other experimental results, Eggler⁴ has suggested that highly alkalic, dense hydrous fluids, similar in composition to that of the micro-inclusions, may exist at high pressures and that the miscibility gap between hydrous fluids and melts may be narrowed or closed in alkalic systems.

The high content of water and carbonate suggests that other volatile components are also enriched. Chlorine levels of 1–3 wt%, reported here, are indeed much higher than those of kimberlites (<0.1%). High levels of rare gases may also be expected and may explain the high ⁴⁰Ar/K ratios and the corresponding excessive apparent ages (6,000 Myr) observed in similar cubic diamonds from Zaire^{23,24}.

Many mantle-derived rocks document the effects of open-system interaction with fluids rich in incompatible elements (mantle metasomatism). The nature of the fluid is still debated^{3,4}, and in most cases there has been no direct observation of such a fluid. The high volatile content of the micro-inclusions, coupled with strong enrichment in K, Na, P, Ti and incompatible trace elements, suggests that the trapped material reported here may represent an effective metasomatizing agent.

Diamond formation

The similarity in composition of micro-inclusions in the coats of coated diamonds and in cubic diamonds is consistent with a common growth environment for the coats and for cubic diamonds, as suggested by the clustering of carbon isotope ratios from the two diamond types^{25,26}. The isotope composition of the cores of the coated diamonds show no correlation with that of the coats^{25–27}, implying different growth conditions for the cores and coats.

The clear compositional distinction between the micro-inclusions and their host kimberlites²⁸ strongly suggest a xenocrystic relationship between the two. Thus, our data do not

support the suggestion of Boyd *et al.*²⁶ that the cubic diamonds are phenocrysts in the kimberlitic melts that transported them to the surface. Rather, we suggest that the diamonds grew from the fluids represented by the micro-inclusions that they trapped.

The broad compositional resemblance of micro-inclusions and kimberlites may reflect a common origin at depth. As in the case of metasomatic fluids and alkaline melts, the identification of one phase as a precursor and the other as a reaction product is difficult, perhaps even impossible. Crystallization of kimberlitic or lamproitic melts at depth has been suggested as the origin of megacrysts²⁹ and of MARID xenoliths³⁰ found in kimberlites. During the final stages of such crystallization, the residue may evolve into a volatile- and incompatible-rich magma, or a supercritical fluid phase may separate^{30,31}. If carbon saturation is also reached, diamonds may precipitate as well and could then trap those fluids or melts during their growth.

Two possible scenarios for the growth and emplacement of micro-inclusion-bearing diamonds have a number of common features and may have included the following events. A parental magma of kimberlitic affinity containing diamond xenocrysts (the cores) is intruded at depth into upper-mantle rocks. Towards the end of crystallization, a residual fluid, enriched in incompatible and volatile elements, would be produced. Growth of coats and cubic diamonds would trap the evolved residue as micro-inclusions. Metasomatism may occur through interaction between the residual fluid and adjacent mantle wall rocks. A later kimberlitic magma then transports the diamonds and the surrounding material to the surface. It seems likely that a requirement for a successful kimberlitic eruption is the presence of volatile-rich residues, left by earlier, unsuccessful events, which act to recharge the final kimberlite with enough volatiles to induce a violent ascent and eruption.

Alternatively, a lithic source, parental to kimberlitic rocks, undergoes incipient melting with the production of a volatile-rich fluid. This magma is intruded in a disruptive event during which growth of cubic diamonds and overgrowths on pre-existing cores (xenocrysts) occurs. This late-stage diamond growth possibly takes place rather rapidly and traps some of the volatile-rich fluid. The cores of coated diamonds represent an earlier stage of long-term growth in different mantle environments. Further melting of the parental lithic reservoir produces a kimberlitic magma which samples the surrounding rocks and may succeed in making its way to the surface, erupting as a kimberlite.

Richardson *et al.*³² suggested that peridotitic garnet inclusions in diamonds equilibrated with an asthenosphere-derived melt rich in alkali, light rare-earth elements and CO₂. They further suggested that the melt remained liquid until diamond crystallization began. It was later argued³³, however, that the strong enrichment of light rare-earth elements in these garnets is too high to be explained by equilibrium crystallization from a silicate melt. The micro-inclusions in the diamonds we studied exhibit strongly fractionated rare-earth patterns. Chondrite-normalized Ce/Eu ratios of 10–21 indicate steeper patterns than those of kimberlites (7–10)³⁴. Thus, the rare-earth patterns of the garnet inclusions could be explained by equilibrium crystallization from fluids similar in composition to the micro-inclusions.

If peridotitic inclusions in diamonds did equilibrate with fluids of this composition, then the partitioning behaviour of major elements between crystals and fluid must deviate significantly from the behaviour in dry basaltic systems. For example, using a typical peridotitic olivine inclusion with Mg/(Fe + Mg) = 0.93 (ref. 19) and the average composition of the micro-inclusions, we obtain (Fe/Mg)_{olivine}/(Fe/Mg)_{fluid} = 0.07, compared with a value of 0.3–0.4 in melting experiments using a dry peridotite³⁵.

Could micro-inclusion-bearing diamonds and diamonds carrying peridotitic inclusions be derived from the same source? Carbon isotope composition of coats and of cubic diamonds ($\delta^{13}\text{C} = -6$ to -8 (ref. 26)) is in the same range as that found

for the majority of peridotitic diamonds ($\delta^{13}\text{C} = -3$ to -8 (ref. 25). This similarity, as well as the strong incompatible-element enrichment of the inclusions in both types of diamonds, are consistent with this possibility. But although peridotitic inclusions and micro-inclusions have been found in separate Zairian diamonds, no diamond containing both types of inclusions has been reported. In addition, the peridotitic garnet inclusions studied by Richardson *et al.*³³ are much older than their host kimberlites, whereas evidence suggests a young age for the coated diamonds. The yellow colour of many coats of coated diamonds is attributed to the presence of single nitrogen centres⁹. Experimental study³⁶ has shown that at mantle temperatures, single nitrogen atoms would combine over a geologically short period, suggesting that the coats are close in age to their host kimberlites. Searches for macroscopic inclusions in

micro-inclusion-bearing diamonds, isotope studies of the micro-inclusions and additional experimental data on potassium-rich systems will help clarify the relations of the two suites of diamond inclusions.

We thank J. T. Armstrong for his essential help in obtaining the EPMA and SEM results, H. Palme and B. Spettel for the INAA analyses, C. T. Pillinger for sharing his samples, data and enthusiasm, Ph. Chevalier of Sibeka, J. Robey of De Beers, J. W. Harris, J. Citron and J. Borden for diamonds, and D. Bell, J. Blum, G. Mattioli, E. Stolper, P. Wyllie, S. Haggerty and E. Roedder for very useful comments and reviews. A preliminary report of this work was presented earlier⁴⁰. Funding was provided by NSF and NASA. O.N., I.D.H. and G.J.W. are at the Lunatic Asylum of the Charles Arms Laboratory (No. 4646/625).

Received 20 July; accepted 19 September 1988.

- Harte, B., Gurney, J. J. & Harris, J. W. *Contr. Miner. Petrol.* **72**, 181–190 (1980).
- Haggerty, S. E. *Nature* **320**, 34–38 (1986).
- Erlank, A. J. *et al.* in *Mantle Metasomatism* (eds Menzies, M. & Hawkesworth, C. J.) 221–311 (Academic, London, 1987).
- Eggler, D. H. in *Mantle Metasomatism* (eds Menzies, M. & Hawkesworth, C. J.) 21–41 (Academic, London, 1987).
- Roedder, E. *Rev. Miner. Vol.* 12 (Miner. Soc. Am., Washington, 1984).
- Plotnikova, S. P., Dudenkov, Yu. A., Malan'ina, R. V. & Kulakov, V. M. *Sov. Phys. Crystal.* **30**, 663–665 (1985).
- Chrenko, R. M., Tuft, R. E. & Strong, H. M. *Nature* **270**, 141–144 (1973).
- Bibby, D. M. *Geochim. cosmochim. Acta* **43**, 415–423 (1979).
- Orlov, Yu. L. *The mineralogy of the diamond* (Wiley, New York, 1977).
- Moore, M. & Lang, A. R. *Phil. Mag.* **26**, 1313–1325 (1972).
- Lang, A. R. in *The Properties of Diamond* (ed. Field, J. E.) 425–469 (Academic Press, London, 1979).
- Custers, J. F. H. *Am. Miner.* **35**, 51–58 (1950).
- Miller, G. H. & Rossman, G. R. *Phys. Chem. Miner.* **14**, 461–472 (1987).
- Armstrong, J. T. in *Microbeam Analysis* (ed. Newbury, D. E.) 239–246 (San Francisco Press, 1988).
- Hunke, J. C., Armstrong, J. T. & Wasserburg, G. J. *Geochim. cosmochim. Acta* **47**, 1635–1650 (1983).
- Davies, G. *Nature* **290**, 40–41 (1981); Davies, G. *Chem. phys. Carbon*, **13**, 1–147 (1977).
- Davies, G., Collins, A. T. & Spear, P. *Solid St. Commun.* **49**, 433–436 (1984).
- Lang, A. R. & Walmsley, J. C. *Phys. Chem. Miner.* **9**, 6–8 (1983).

- Meyer, H. O. A. in *Mantle Xenoliths* (ed. Nixon, P. H.) 501–522 (Wiley, New York, 1987).
- Kresten, P. in *Lesotho Kimberlites* (ed. Nixon, P. H.) 269–279 (1973).
- Ryabchikov, I. D. & Boettcher, A. L. *Am. Miner.* **65**, 915–919 (1980).
- Ellis, D. E. & Wyllie, P. J. *Am. Miner.* **65**, 540–556 (1980).
- Zashu, S., Ozima, M. & Nitoh, O. *Nature* **323**, 710–712 (1986).
- Podosek, F. A., Pier, J., Nitoh, O., Zashu, S. & Ozima, M. *Nature* **334**, 607 (1988).
- Galimov, E. M. *Geochimiya* **8**, 1091–1118 (1984).
- Boyd, S. R. *et al. Earth planet. Sci. Lett.* **86**, 341–355 (1987).
- Swart, K., Pillinger, C. T., Millard, H. J. & Seal, M. *Nature* **303**, 793–795 (1983).
- Fieremans, M., Hertogen, J. & Demaiffe, D. in *Kimberlites I* (ed. Kornprobst, J.) 107–120 (Elsevier, Amsterdam, 1984).
- Schulze, D. J. in *Mantle Xenoliths* (ed. Nixon, P. H.) 432–451 (Wiley, New York, 1987).
- Waters, F. G. *Contr. Miner. Petrol.* **95**, 523–533 (1987).
- Boyd, S. R., Nixon, P. H. & Boctor, N. Z. *Contr. Miner. Petrol.* **86**, 119–130 (1984).
- Richardson, S. H., Gurney, J. J., Erlank, A. J. & Harris, J. W. *Nature* **310**, 198–202 (1984).
- Shimizu, N. & Richardson, S. R. *Geochim. cosmochim. Acta* **51**, 755–758 (1987).
- Mitchell, R. H. *Kimberlites* (Plenum, New York, 1986).
- Takahashi, E. & Kushi, I. *Am. Miner.* **68**, 859–879 (1983).
- Evans, T. & Qi, Z. *Proc. R. Soc. Lond. A* **381**, 159–178 (1982).
- Thompson, W. K. *Trans. Faraday Soc.* **61**, 2635–2640 (1965).
- Smith, C. B., Gurney, J. J., Skinner, E. M. W., Clement, C. R. & Ebrahim, N. *Trans. geol. Soc. S. Afr.* **88**, 267–280 (1985).
- Bergman, S. C. in *Alkaline Igneous Rocks* (eds Fitton, J. G. & Apton, B. G. J.) 103–190 (Geol. Soc. spec. Publ. no. 30, London, 1987).
- Navon, O., Hutcheon, I. D., Rossman, G. R. & Wasserburg, G. J. *Lunar planet. Sci.* **XIX**, 827–828 (Lunar planet. Inst., Houston, 1988).

Structure of a phage 434 Cro/DNA complex

Cynthia Wolberger*, Yicheng Dong*, Mark Ptashne & Stephen C. Harrison†

Department of Biochemistry and Molecular Biology, Harvard University and †Howard Hughes Medical Institute, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA

Comparison of the crystal structure of a complex of the phage 434 Cro protein and a synthetic DNA operator with the complex of the same operator and the 434 repressor DNA-binding domain shows different DNA conformations in the two structures. Binding of the protein determines the precise conformation of the DNA in each case.

THE recognition of specific DNA sequences by proteins containing a helix-turn-helix element¹ depends on contacts between this common structural motif and DNA. The first crystal structure of such a protein bound to its operator site, that of the DNA-binding domain of 434 repressor (R1-69) in complex with a 14-base pair oligomer², showed that the helix-turn-helix is indeed positioned to be in contact with DNA bases and the sugar-phosphate backbone. It also showed that the DNA in the complex is bent, with significant variation in twist from one base-pair step to another, suggesting that non-contacted base pairs can influence repressor affinity by determining how readily DNA adopts the required conformation. The essential features observed in this structure have been confirmed by the recent crystallographic study at higher resolution of R1-69 bound to a 20-base pair DNA fragment³. This structure has also shown a network of bifurcated hydrogen bonds among non-coplanar base pairs in the operator which helps explain why certain

sequences of non-contacted bases are more favourable than others for repressor binding.

The Cro and repressor proteins of bacteriophage 434 together regulate the switch between lytic and lysogenic growth of the phage⁴. These proteins are far more similar in structure and sequence than the corresponding proteins of bacteriophage lambda, which are structurally congruent only within the helix-turn-helix motif⁵. The transcriptional regulation depends on the ability of the two proteins to bind with different orders of affinity to the same set of six partially symmetric, 14-base pair operator sites. Of the twelve half sites, eleven have identical 5'-ACAA sequences in their first four positions; the twelfth, one side of O_R3, has ACAG. Studies using systematic base substitutions in a 'standard' operator sequence show that the G-for-A substitution at position 4 in O_R3 has a significant unfavourable effect on repressor binding, but only slightly reduces the binding of Cro (R.P. Wharton, G.B. Koudelka and M.P., unpublished results). Thus, two remarkably similar proteins are affected differently by a single substitution in otherwise identical sequences.

* Present addresses: University of California at San Francisco, Department of Biochemistry and Biophysics, San Francisco, California 94143-0448, USA (C.W.); Institute of Biophysics, Academia Sinica, Beijing, People's Republic of China (Y.D.).

Table 1 Data statistics

Combination statistics			
Data set	Resolution	Number of reflections	R_{sym}
Native (film)	3.2	14,143	0.144*
Native (detector)	7.0	3,782	0.086
Derivative (detector)	7.0	4,050	0.075
Mean isomorphous difference = $\sum F_{\text{nat}} - F_{\text{der}} / \sum F_{\text{nat}} = 0.13$			
$R_{\text{sym}} = \sum \sum I_{ij} - \bar{I}_i / \sum \sum I_{ij}$			
Percentage of data present in merged native data set			
Resolution range (Å)	Number of reflections	% Possible reflections†	
20.0–10.5	1,192	98	
10.5–7.0	2,825	92	
7.0–6.0	1,815	72	
6.0–5.5	1,315	63	
5.5–5.0	1,300	61	
5.0–4.5	1,437	66	
4.5–4.0	1,644	65	
4.0–3.5	1,746	57	
3.5–3.2	897	48	
Total	14,353	69	

* R_{sym} on all data; whole spot $R_{\text{sym}} = 0.113$.

† The percentage of possible reflections was calculated by considering only those reflections that fell within the limits of anisotropic diffraction defined by a cylinder along the (1 0 5) direction, of length 1/3.2 Å and radius 1/5.5 Å. This cylinder represents 46% of the full sphere at 3.2 Å resolution.

Data were collected on 2–3° oscillation photographs, recorded on CEA Reflex 25 film, using the Cornell High Energy Synchrotron Source (CHESS). The films were scanned with an Optronics P1000 film scanner on a 50 µm raster, and integrated intensities were obtained with the program SCANFILM (ref. 30). Further data processing and scaling were carried out with programs from P.R. Evans (MRC Laboratory of Molecular Biology, Cambridge, UK). Partially observed reflections that were at least 60% recorded were corrected to their fully recorded equivalents³¹. Area detector data were collected using a Xentronics imaging proportional counter with CuK α radiation from an Eliot GX-13 rotating anode X-ray generator. These data were recorded in 10° oscillation frames, processed as described⁹, and scaled to the recorded data on film by dividing the frames into 10-frame batches. The final merged native data set contained both the film data and the area detector data. Data from derivative crystals were collected on the area detector, processed as described above, and scaled to the merged native data set containing data from both area detector and film.

In order to compare how the two proteins bind to DNA, we undertook the X-ray crystallographic study of 434 Cro bound to the same 14-base pair operator present in the R1-69/14-mer crystals. As Cro and R1-69 are so similar in sequence and in three-dimensional structure^{6,7}, we anticipated that they would form very similar complexes when bound to the same DNA fragment. Our surprising finding is that the same 14-base pair operator adopts a different conformation, depending on whether Cro or R1-69 is bound. The relations between the residues of the helix-turn-helix units and the DNA bases are therefore different in the two complexes, and the differences have implications for operator recognition.

The structure of the 434 Cro-DNA complex reported here has been determined from crystals that diffract to 3.2 Å in one direction and to 5.5 Å in perpendicular directions. Like R1-69, Cro binds as a dimer to B-form DNA, with each recognition helix lying in the major groove of the DNA. With reference to the DNA in complex with R1-69, the DNA in the Cro complex is more uniformly overwound and significantly less bent. There is a corresponding difference in the relative orientation of the two protein monomers in each structure. Moreover, as a result

of the DNA conformational differences, interactions between the turn in the helix-turn-helix unit and the sugar-phosphate backbone of DNA are not the same. In short, differences between the structures are not confined to the vicinity of base pair 4, and a model explaining why Cro and R1-69 respond differently to substitution at this position must take these distributed differences into account.

Structure determination

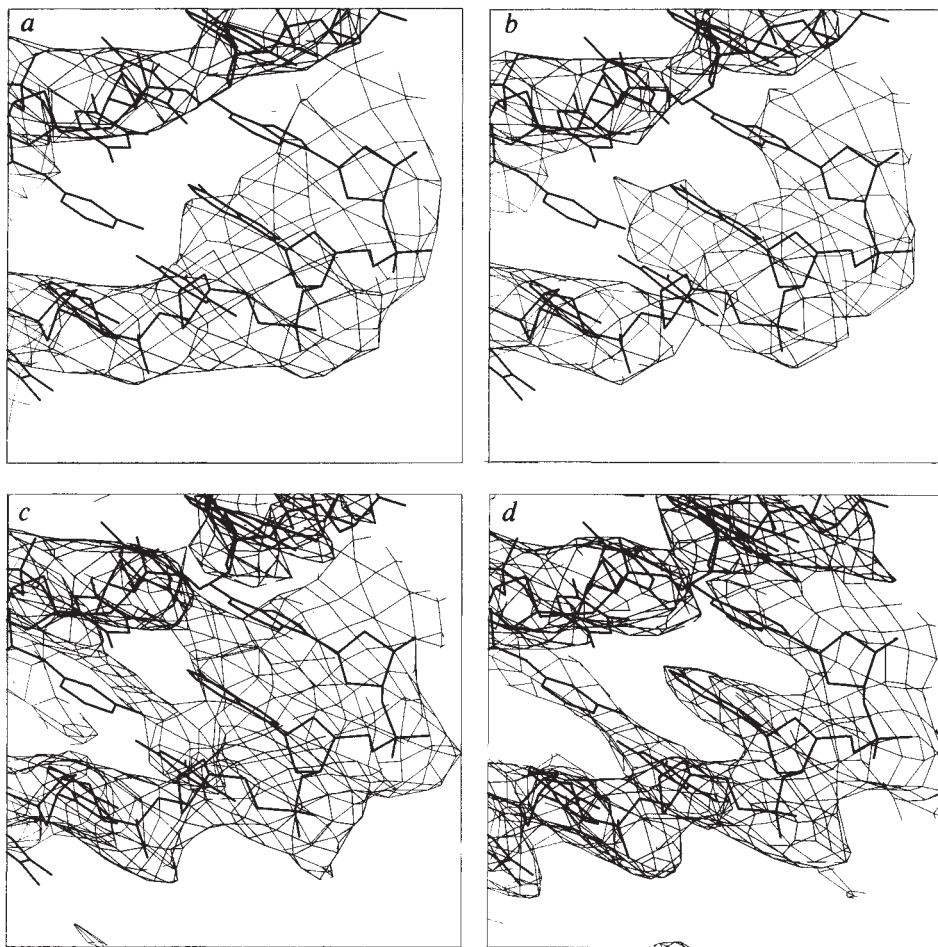
Crystals of the complex of 434 Cro and the synthetic 14-base pair operator were prepared as described previously⁸. Diffraction from the crystals is anisotropic, extending to 3.2 Å along one direction and to 5.5 Å in perpendicular directions. The space group is C2, with $a = 351.8$ Å, $b = 78.2$ Å, $c = 129.1$ Å, $\beta = 129.4$ degrees. Each strand of the DNA has the sequences 5'-OH-d(A-C-A-A-T-A-T-A-T-A-T-T-G-T)-OH3'. Isomorphous derivative crystals were prepared using a DNA oligomer that was identical except for the substitution of 5-bromodeoxyuridine for dT at position 7. Native and derivative data to 7 Å resolution were collected using a Xentronics imaging proportional counter⁹ mounted on an Eliot GX-13 rotating anode X-ray generator with CuK α radiation. Data to 3.2 Å resolution were collected on film from native crystals using the synchrotron radiation at the Cornell High Energy Synchrotron Radiation Source (CHESS). Combination statistics are summarized in Table 1.

Heavy atom positions were located using phases from a molecular-replacement model that was constructed from the coordinates of the 434 R1-69/14-mer (ref. 2) and the packing in the Cro/DNA complex crystals⁸. The model for the Cro/DNA complex was constructed by replacing each monomer of R1-69 with the coordinates of the uncomplexed Cro protein, determined at the resolution of 2.35 Å (ref. 10). All main-chain atoms and side-chain β -carbon atoms were included. The complexes were then arranged to stack end-to-end with 11₃ screw symmetry, forming long rods extending along the (5 0 1) direction. A total of 5.5 complexes, or 11 half complexes, were in the asymmetric unit. This model was subjected to rigid-body refinement using the program CORELS (ref. 11). Each Cro monomer and DNA half-operator were treated as separate domains, and positions of the 22 rigid bodies were refined against data from 15 to 7 Å. The R -factor on model structure factors dropped from 0.50 to 0.28. In the refined model, adjacent oligomers were stacked end-to-end, but the strict 11₃ symmetry was broken.

Phases from the refined molecular-replacement model were used to calculate a 7 Å heavy-atom difference Fourier map. The 11 bromine atoms appeared clearly as strong peaks. Heavy atom positions determined from this map were refined by conventional methods¹² to a final figure of merit of 0.35. Single isomorphous replacement (SIR) phases were calculated from these bromine positions and used to produce a 7 Å electron density map. The map showed a clear distinction between regions of solvent and structure and some features that resembled double-helical DNA. The set of transformations describing the positions of the 11 half complexes were determined from the refined molecular replacement model and used to average this SIR map¹³. The averaged SIR map showed clearly the 5 α helices of each Cro monomer and the DNA backbone.

The 7 Å phases were refined by non-crystallographic symmetry phased refinement^{13–15}. The envelopes required for the phase-refinement calculation were generated from the refined molecular replacement model described above. Electron density maps were averaged over the 11 copies of the half-complex in the asymmetric unit, using the transformations derived from the refined model. New phases were calculated by Fourier transformation of the averaged maps and used to initiate the next cycle of refinement. The calculation converged after 7 cycles to an R -factor of 0.221 and a correlation coefficient of 0.918. In the final 7 Å map, the density corresponding to the Cro helices and the sugar-phosphate backbone of the DNA were completely continuous.

Fig. 1 Part of the electron density map at different stages of the structure determination, showing a portion of the DNA sugar-phosphate backbone. The coordinates of the final model are superimposed on all maps. *a*, Map at 7 Å resolution, following refinement of SIR phases by non-crystallographic symmetry averaging. *b*, Map obtained by phase extension from 7 to 5.2 Å. Phosphate groups begin to appear as bulges and some density is found in the base-pair region. *c*, Final map obtained by phase extension to 3.2 Å. Phosphate groups are well-defined and density is found throughout the base-pair region, although the bases are not well resolved. *d*, $2F_o - F_c$ map at 3.2 Å.



As no useful heavy atom-derivative data beyond 7 Å resolution were available, phase extension based on non-crystallographic symmetry¹⁶⁻¹⁸ was used to determine phases to 3.2 Å. This method was preferred to the use of molecular replacement phases, which would have introduced greater bias into the model. In order to determine the averaging transformations to be used in the phase extension, the molecular replacement model

was further refined against data to 3.2 Å. In this second round of refinement the protein side chains were included in the model, but their conformations were not altered from those found in the free protein. Rigid-body refinement was carried out against data to 3.2 Å, and new averaging transformations were determined. The phase extension was possible with anisotropic resolution, because the axis of non-crystallographic symmetry coincided with the axis of anisotropy.

Beginning with the final averaged 7 Å map described above, the resolution of the structure determination was extended in steps of $1/d = 0.01 \text{ Å}^{-1}$. At each resolution step, the map was averaged for several cycles until convergence was achieved, as reflected in a small overall phase change per cycle (less than 2 degrees). Missing structure factors were substituted with calculated values from the inverted map^{19,20}. Because of the anisotropic nature of diffraction from the Cro-DNA crystals, only those structure factors were substituted that fell within the region of reciprocal space where diffraction was expected to occur. This region was defined by a cylinder $1/5.5 \text{ Å}$ in radius whose axis was coincident with that of the DNA-protein rods. The phase extension from 7 to 3.2 Å was carried out in a total of 17 resolution steps. The final 3.2 Å map had an averaging *R*-factor of 0.27. A portion of the electron density map at different stages of the structure determination is shown in Fig. 1.

The electron density map showed clear and continuous density for both protein and DNA. Whereas the 7 Å map had exhibited a smooth surface for the DNA sugar-phosphate backbone, the electron density map calculated by phase extension had distinct bulges corresponding to phosphate groups and the backbone had a staircase-like appearance. Density was present in the base-pair region, but individual bases were not resolved. The five Cro α helices appeared as grooved rods with bulges at the positions of most side chains. Density corresponding to complete

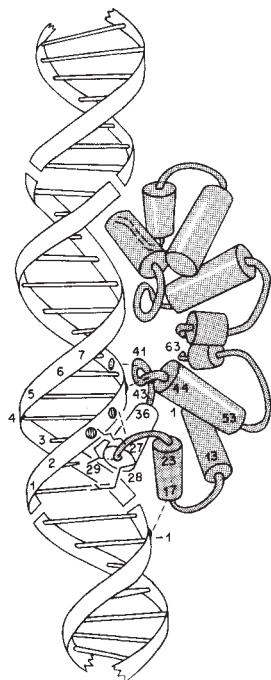


Fig. 2 General view of 434 Cro/14-mer complex. Parts of adjacent 14-mers are shown, stacked as in the crystal. Residue numbers are given at the beginning and ends of the five α helices. In addition, certain residues important for the interaction with operator are shown explicitly. Base pairs in one half-site are numbered (1-7) along the backbone ribbon in a 5' to 3' direction.

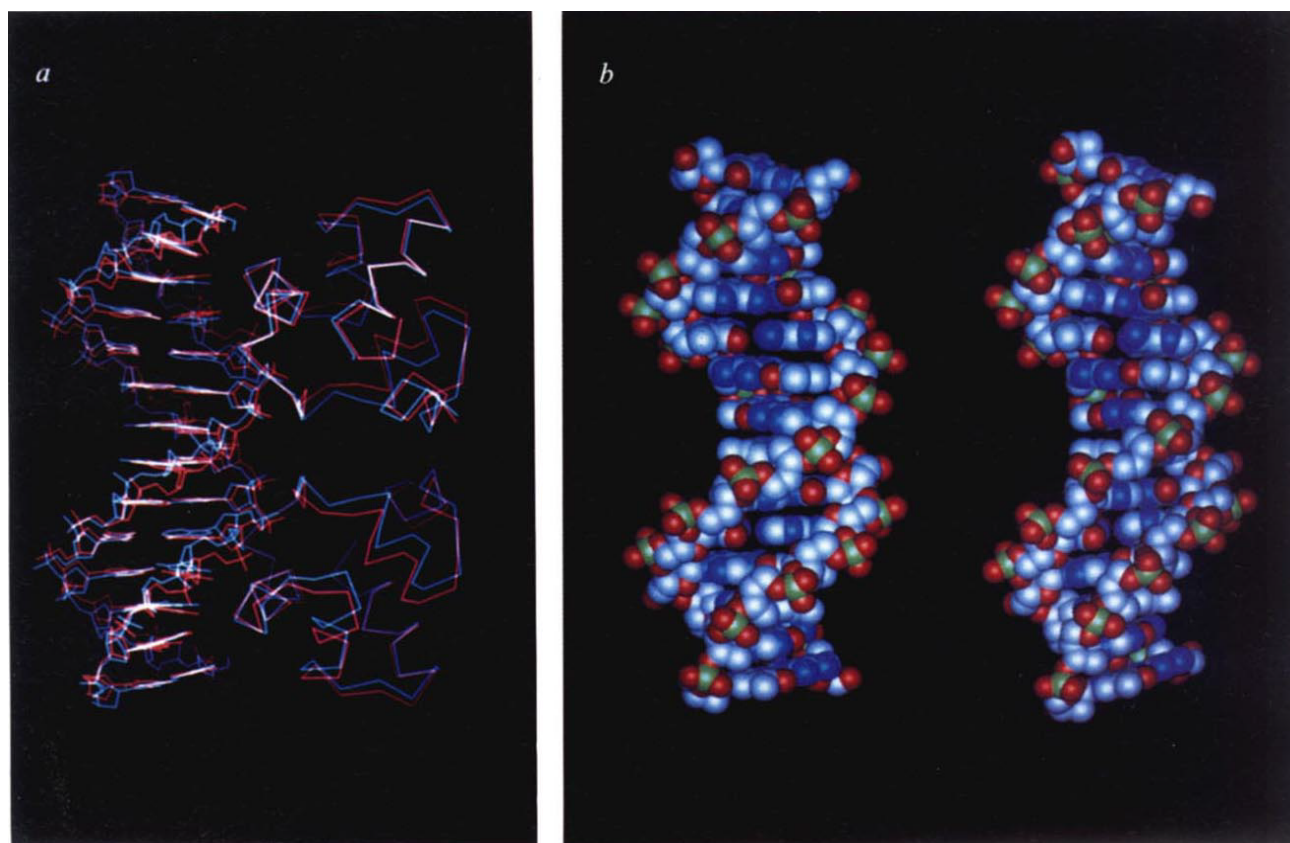


Fig. 3 *a*, Comparison of the Cro/DNA and R1-69/DNA complexes. The Cro (red) and R1-69 (blue) complexes were superposed by a least-squares procedure to produce the best fit between the protein dimers. The view is perpendicular to the dyad, and only the α -carbon atoms of the proteins are shown. The Cro monomers are further apart from one another than the R1-69 monomers. *b*, Space-filling models of the DNA in the Cro and R1-69 complexes. The DNA in the Cro complex (left) is essentially straight, but that in complex with R1-69 (right) is bent at the ends.

side chains appeared for only a few residues. A model for the DNA was built by fitting coordinates for idealized planar nucleotide base pairs to the electron density, guided by the positions of the phosphate groups and the appearance of the DNA backbone density. The DNA model was regularized using the least-squares refinement program CORELS (ref. 11) to impose correct geometry.

The atomic coordinates for Cro, determined from crystals of the free protein, were fitted to the electron density corresponding to the protein²¹. As side chains were not well-defined in the map, we did not reposition them from their conformations in the free protein, except as indicated in the following discussion on model-building. Limited refinement of the model was carried out against data that had been scaled to the model structure factors by application of an anisotropic temperature factor²². The *R*-factor on the model presented is 36.6%.

Overview of the Cro-DNA complex

The model of the Cro-DNA complex is shown in Fig. 2. The DNA is in the B-form, and the Cro dimer lies with its twofold axis superposed on that of the DNA operator. Each monomer consists of 5 α helices, α 1–5. Cro residues were given the same numbers as the homologous residues in R1-69; the first two Cro residues are numbered –1 and 0. Residues 16–36 form the helix-turn-helix motif. In each monomer the recognition helix, α 3 (residues 28–36), lies in the major groove. The amino terminus of α 2 (residues 16–22) lies against the sugar-phosphate backbone on one side of the major groove, and the loop of extended chain connecting α 3 and α 4 (residues 37–44) lies against the sugar-phosphate backbone on the other side of the groove. Helix 4 is positioned with its amino terminus near a phosphate and

its carboxyl terminus pointed away from DNA. The final short helix, α 5, lies between the end of α 4 and the amino terminus of α 1. No density is seen in the map corresponding to the 6 carboxyl-terminal residues, which are also disordered in crystals of the free protein¹⁰. Within the limits of the resolution of our map, there is no apparent conformational change of the Cro monomers on binding to DNA.

Comparison of the Cro and R1-69 complexes

A superposition of the Cro and R1-69 (ref. 2) complexes reveals a fundamental similarity, but there are also important differences between the two structures in the conformations of both DNA and protein dimer. These differences can be divided into three categories: (1) the relative disposition of the two subunits in each dimer, (2) the overall conformation of the two operators, and (3) the relation between protein and DNA in each operator half-site. We discuss the details of the conformation of the Cro complex and compare them with those of the R1-69 complex in the sections that follow.

Dimer conformations. Comparison of the Cro and R1-69 complexes, as shown in Fig. 3*a*, demonstrates that the two protein monomers in each structure have rather different relative orientations. With respect to R1-69 positions, each Cro monomer appears to be rotated about the axis of α 3 in such a way as to approximately preserve the spacing between the two α 3 helices while widening the gap across the dimer interface. The magnitude of the difference is apparent when least-squares methods are used to superimpose only one monomer in each structure, leaving the second monomer unconstrained. Because of differences in the quaternary organization of the two dimers, this superposition gives poor alignment of the second monomers (Fig. 4).

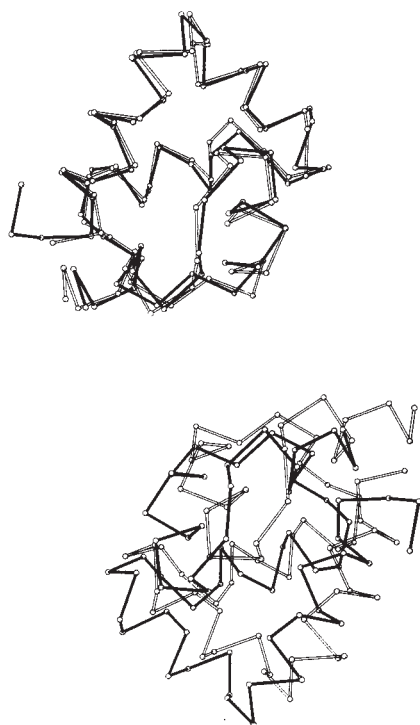


Fig. 4 Superposition of Cro and R1-69 dimers showing the differences in quaternary structure. The α -carbon backbone of the Cro (black lines) and R1-69 (open lines) dimers are shown. Least-squares fitting of α -carbon coordinates was used to superpose one half of the R1-69 dimer on one half of the Cro dimer. The protein monomers (upper half) superpose quite well, but the relation to the opposing monomer (lower half) is different in the Cro and R1-69 dimers. The view is approximately along the dyad from 'behind' the protein: that is, DNA would be buried in the page.

Positions of corresponding alpha carbon atoms in the superposed monomer have a root-mean-square difference of 0.77 Å, whereas α -carbon atom positions in the other monomer have a root-mean-square difference of 3.3 Å.

The difference in the relative disposition of protein monomers in the Cro and R1-69 complexes comes from different protein-protein contacts across the dimer interface. This interface in the Cro complex contains two bulky side chains: Phe 46 in $\alpha 4$, and Tyr 61 in $\alpha 5$. The map clearly shows the position of the Tyr 61 ring. Phe 46 can make van der Waals contact with Tyr 61 in the opposing monomer, giving rise to a pair of aromatic ring interactions. In the R1-69/14-mer complex, the corresponding packing of the smaller residues Pro 46 and Leu 61 permits closer approach of $\alpha 4$ and $\alpha 5$.

DNA conformation. The same 14-base pair DNA oligomer adopts a different conformation, depending on the protein to which it is bound. In brief, the DNA in complex with Cro is straight and uniformly overwound, with no large variation in the width of the minor or major grooves, whereas the DNA in complex with R1-69 is bent, overwound near its centre and underwound near its ends, with striking variation in minor groove width (Fig. 3b). The DNA forms a B-type helix in both cases. The operator bound to Cro has an average of 10.1 base pairs per turn and an additional 17 degrees of twist, compared with the same operator bound to R1-69. This explains in part why the Cro complex crystallizes differently from the R1-69 complex. The Cro/DNA complexes stack end-to-end in the crystals to form pseudocontinuous helices with an average angular displacement per complex of 130.9° (ref. 8), whereas the complexes in the R1-69 crystals stack with an angular displacement of 120° (ref. 23).

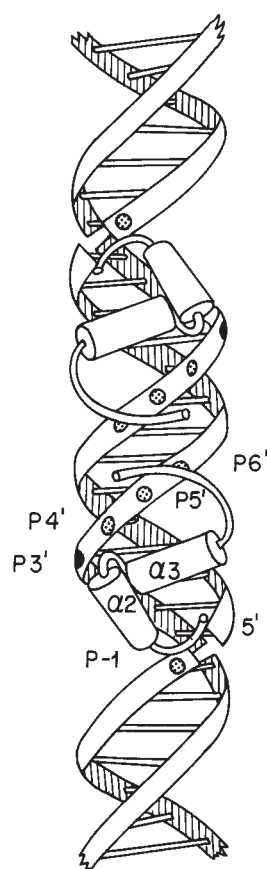


Fig. 5 Contacts between 434 Cro and DNA backbone phosphates. The numbering scheme is:

-2'	-1'	1	2	3	4	5	6	7	7'	6'	5'	4'	3'	2'	1'
G	T	A	C	A	A	T	A	T	A	T	A	T	T	G	T
0															
C	A	T	G	T	T	A	T	A	T	A	T	A	A	C	A
-2	-1	1'	2'	3'	4'	5'	6'	7'	7	6	5	4	3	2	1

Phosphates 1 and 1' are missing as the DNA fragments have 3' and 5'-OH groups. The end-to-end stacking of 14-mers in the crystals is implied by the numbering -1' and so forth. Phosphate 6' appears to have hydrogen bonds from peptide -NH groups 40 and 41; phosphate 5', from -NH 43; phosphate 4', from -NH 27; phosphate 3', from Lys 27 (ϵ -NH₃⁺); phosphate -1', from -NH 17.

In the 434 Cro/DNA complex, there is only a slight bending of the DNA, but the DNA in the R1-69/14-mer complex is bent to a radius of curvature of less than 100 Å (Fig. 3b). This difference in bending is consistent with the results of gel mobility assays, which indicate that binding of 434 repressor or of R1-69 to the 14-base pair operator bends the DNA, whereas binding of Cro does not²⁴.

Protein interactions with the sugar-phosphate backbone. Each Cro monomer interacts in its half-site with at least five phosphates (Fig. 5), corresponding to those identified in ethylation interference experiments²⁵. With the exception of one (P3'), ethylation of the same phosphates interferes with repressor binding²⁶. The contacts to P6', P5' and P-1 (see Fig. 5 for numbering scheme) are similar to those in the R1-69 complex. P6' lies within hydrogen-bonding distance of the main-chain -NH groups of Lys 40 and Arg 41, and P5' lies within hydrogen-bonding distance of the main-chain -NH of Arg 43. The side chain of Arg 43 projects into the minor groove and it can contact phosphates 6' or 5' on either side. At P-1, an important interaction occurs between stacked complexes in the crystal: the peptide -NH of Gln 17, located at the N-terminus of $\alpha 2$, and the side chain of Thr 16

approach this phosphate closely enough to form hydrogen bonds. This interaction probably mimics one made in continuous DNA (refs 2, 3 and 26).

The most significant difference in the way Cro and R1-69 interact with the DNA backbone is at the turn between helices 2 and 3. In the Cro/DNA complex, the turn lies close to the DNA backbone, permitting contacts between phosphates P4' and 3' and Lys 27. The peptide bond NH is in a position to form a hydrogen bond with P4', and we can build the positively charged lysine side chain so that it forms a salt bridge with P3'. These contacts are not present in the R1-69 complex because the DNA backbone between P4' and P3' lies 2.5 Å further away from the $\alpha 2/\alpha 3$ turn of R1-69 (Figs 5 and 6), and because residue 27 is Thr, not Lys. The backbone shift in the R1-69/14-mer complex is probably a consequence of the way in which DNA is bent by the protein dimer. This shift moves P4' out of range of N27. Moreover, the branched Thr 27 side chain of R1-69, which cannot reach to P3', would also interfere with an N27-P4' hydrogen bond. We note that this significant local difference between the two structures occurs precisely at the nucleotide (4') at which specificity is different. Its consequence for protein/base-pair interactions is described in the next section.

Four of the five contacted phosphates can form hydrogen bonds with peptide —NH groups. In two cases, the —NH groups lie at the amino terminus of an α helix. These characteristics of protein-DNA-backbone interactions are also observed in the R1-69/14-mer and 20-mer complexes^{2,3}.

Base-pair contacts. The limited resolution of the map permits only an approximate description of interactions with base pairs, because clear indication of side-chain conformation is not generally present. Three Gln residues and a Leu—residues 28, 29, 32 and 33, all in $\alpha 3$ —are in positions to project into the major groove and to contact the outer four base pairs in each operator half-site. In the conformations found in the free protein, Gln 28 and Gln 29 would clash with parts of the outer two base pairs of the operator, and a refolding of those side chains must therefore occur in the complex. These residues are present in homologous positions in R1-69 and contact base pairs 1 and 2 respectively. We find that their side-chain conformations can be adjusted so that hydrogen bonds form between Gln 28 and adenine in base pair 1, and between Gln 29 and guanine in base pair 2. The thymine methyl group in base pair 3 is then in van der Waals contact with non-polar parts of the Lys 27 and Gln 29 side chains. These interactions are essentially identical to those made by repressor^{2,3}. They uniquely determine recognition of the cognate sequence (ACA) at positions 1–3.

The thymine methyl group in base pair 4 lies near the backbone carbonyl and C β of Gln 29 and C δ of Leu 33. With respect to the homologous contact in the R1-69 complex, it lies closer to $\alpha 3$, because the corresponding sugar-phosphate backbone lies closer to the $\alpha 2/\alpha 3$ turn (see above and Fig. 6). The tight contact for the thymine methyl group suggests that its removal, by substitution of G–C for A–T at base pair 4, could relieve strain and thus compensate for loss of van der Waals energy. We discuss in the final section how this effect could account for differential discrimination by Cro and repressor. The position of $\alpha 3$ in the major groove rules out any contacts between side chains of helix 3 and base pairs 5, 6 and 7, and there appear to be no other direct interactions between Cro and the major groove of the DNA at these base pairs. Arg 43, which projects into the minor groove, is too far from the central base pairs to form hydrogen bonds. Model building shows that there can be no contact between the six disordered carboxyl-terminal residues and DNA base pairs.

The protein determines operator conformation

The different operator conformations found in the structure described here and in the R1-69/14-mer complex show that interactions with the proteins significantly influence the confor-

mation of the DNA. A comparison of the two structures reveals common features, important in positioning protein on DNA, and differences, which produce the distinct operator conformations. The common features of complexes formed by Cro and R1-69 with the 14-base pair operator are as follows: (1) Residues at the amino terminus of $\alpha 2$ contact sugar-phosphate backbone on one side of the major groove and residues in the loop between $\alpha 3$ and $\alpha 4$ contact backbone on the other side of the groove. Hydrogen bonds between corresponding peptide —NH groups and DNA phosphates appear to be particularly important. These interactions clamp the helix-turn-helix on to the major groove. (2) The recognition helices, $\alpha 3$, are then positioned in the major groove so that side chains can have essentially identical interactions with base pairs 1–3. (3) Corresponding residues in the $\alpha 3/\alpha 4$ loop, in $\alpha 4$, and in $\alpha 5$ are involved in contacts between adjacent monomers across the protein dimer interface.

The difference in the DNA structures is a result of differences in the way Cro and R1-69 interact with the 14-base pair operator. The principal differences between the complexes are: (1) The side chains at the dimer interface differ in size and hence in how they pack. As a result, the relative orientations of monomers in the two complexes are not the same. (2) In a complex with operator, the relative orientation of two Cro monomers is consistent with essentially straight DNA (each monomer clamping onto its half-site as described above). By contrast, the relative orientation of two R1-69 monomers in a complex imposes a bend. (3) In addition to this bending, there is a shift of DNA backbone near phosphates 3' and 4' away from close contact with the $\alpha 2/\alpha 3$ turn. Thus, in the complex of Cro with straight DNA, N27 can form a hydrogen bond with P4' and Lys 27 side chain and a salt bridge with P3'. In the complex of R1-69 with bent DNA, corresponding contacts are absent (see Figs 5 and 6).

Differential discrimination

Repressor and Cro of phage 434 both bind specifically to operator sites containing ACAA at positions 1–4. The only exception is O_R3, which contains ACAG in one half-site. The affinity of Cro for O_R3 is relatively unimpaired by this exceptional sequence, because the 'preference' of Cro for A over G at position 4 is slight. By contrast, binding of 434 repressor is strongly affected by the substitution. The difference appears to be critical in accounting for a central feature of the regulatory switch: of the sites in O_R, Cro binds most tightly to O_R3, and repressor least tightly.

How can comparison of the R1-69 and Cro complexes explain this difference? In the repressor complex, Gln 33 forms a hydrogen bond with thymine in base pair 4. The corresponding residue in Cro is Leu. But the simple notion that the Gln to Leu substitution diminishes specificity at base pair 4 by removing the interaction is inconsistent with experimental results using modified repressors. The single change Gln 33 to Leu in repressor leaves discrimination at base pair 4 unaltered (G. Koudelka and R.P. Wharton, unpublished results). Moreover, the concerted substitution of three residues in the repressor recognition helix for those found in Cro does not lead to a precisely Cro-like affinity order²⁷. Because of the different configurations of the DNA major groove in the Cro and repressor complexes, substitution of one or a few Cro-like residues into 434 repressor cannot mimic the normal approach of these residues to base pairs. That is, Cro and R1-69 monomers present essentially identical contacts with DNA over most of their interacting surfaces, but the straight DNA in one complex and the bent DNA in the other have somewhat different surfaces, with differences concentrated near base pair 4. In the complex with Cro there appears to be a tight contact for the methyl group of thymine 4' that can be relieved either by displacement of the nucleotide, as in the R1-69 complex, or by substitution of G–C for A–T, as in O_R3. In the latter case, elimination of the tight contact may compensate for loss of van der Waals contacts from the methyl group. We propose that this compensation explains

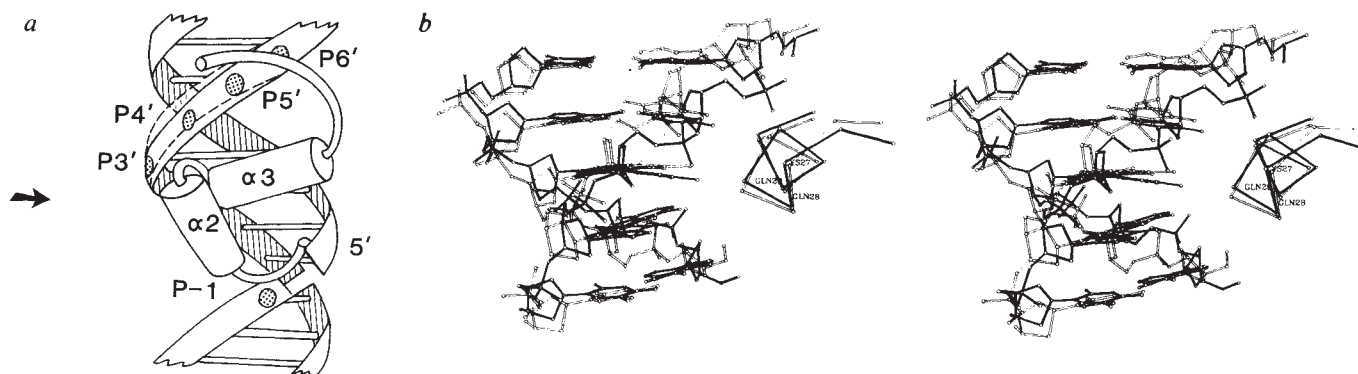


Fig. 6 Relative positions of the N-terminus of $\alpha 3$ and the DNA backbone. *a*, The diagram shows the helix-turn-helix-loop structure in 434 Cro as it clamps on to the major groove of an operator half site. The course of the DNA backbone is represented by the ribbon. In the R1-69 complex, the sugar phosphate backbone follows essentially the same path as in the Cro complex, except in the region near P4', where it follows the path indicated by the dashed lines. This conformational difference lies near the turn between $\alpha 2$ and $\alpha 3$. The arrow shows the overall viewing direction for the stereo drawing. *b*, The stereo view shows the Cro (solid lines) and R1-69 (open lines) complexes, which have been superposed by least-squares alignment of the α -carbons of residues 26-44. Only the α -carbons of residues 25-34 are shown. Lys 27 of Cro is labelled; in R1-69, this residue is threonine. The residue at positions 28 and 29 in both proteins is glutamine. The sugar-phosphate backbone of the DNA approaches the amino-terminus of $\alpha 3$ more closely in the Cro complex than in the R1-69 complex, apparently permitting a hydrogen bond between the $-\text{NH}$ group of Lys 27 and phosphate 4'.

why 434 Cro binds as well to operators with A-T or G-C as to the fourth base pair. We also suggest that the structure cannot readily accommodate a purine at position 4' as there is no hydrogen bond donor for the buried N7, thus accounting for poor binding to operators with T-A or C-G. In short, by imposing distinct conformations on the operator, the two proteins appear to set up distinct alternatives for recognition.

The affinity of 434 Cro and repressor for operator DNA is affected by the central base pairs, which are not contacted by either protein²⁸. The effect of base-pair substitutions at the centre of the operator must therefore be due to their effect on DNA structure. In the R1-69/14-mer and R1-69/20-mer structures, the operators are bent and over-twisted at the centre, with an accompanying narrowing of the minor groove. The results of *in vitro* binding studies using mutant repressors with altered dimer contacts and with wild-type repressor and nicked operators indicate that the non-contacted central base pairs influence repressor binding affinity by affecting the ability of the operator to twist²⁴. Cro does not significantly bend DNA, nor does it compress the minor groove, but it nonetheless exhibits the same sensitivity to central base-pair changes as R1-69. We conclude that the central base pairs likewise influence Cro affinity by affecting the ability of the operator to twist. Indeed, we suppose that some of the network of bifurcated hydrogen bonds seen in the R1-69/20-mer complex³ may also be found in a similar complex with Cro, to the extent that the array of non-coplanar base pairs participating in the network represents a response to twist required at the central base-pair steps rather than to bending and minor-groove compression.

The results presented here taken together with the results of mutagenesis experiments show that recognition of specific sequences cannot be the result of a simple code in which certain residues in the recognition helix confer specificity for certain bases. Early comparison of the structures of uncomplexed lambda repressor and Cro suggested a similar conclusion⁵. The details of the conformation of the DNA and the relative position of the recognition helix in the major groove are not the same in the 434 Cro and R1-69 complexes, and hence the energy of interaction of amino-acid side chains with DNA must depend on the conformational details of the complex. Even where the R1-69 and Cro complexes are very similar, as in the contacts made by Gln 28 and Gln 29 to the conserved outer base pairs, small differences have measurable consequences. When Ala is substituted for Gln 28 in R1-69, it confers on the protein the ability to recognize an operator in which T replaces A at position 1 (ref. 29). However, this same amino-acid substitution does not enable Cro to bind to the same altered operator (R.P. Wharton, unpublished results). Although it may be possible to devise a set of recognition rules for a given DNA-binding protein, they will not generally be true, even for proteins that recognize DNA with the same structural motif.

We thank W. Schildkamp and the staff of CHESS for help with data collection, J. Moulai for assistance with area-detector data processing, and M. Drott for laboratory assistance. We also thank A. Aggarwal, A. Mondragon, G. Koudelka, R. Wharton, J. Anderson, P. Bjorkman, W. Weis and M. Saper for discussion and J. Sussman for programs and advice. This work was supported by the NIH (S.C.H. and M.P.).

Received 23 August; accepted 20 September 1988.

- Pabo, C. O. & Sauer, R. T. *A. Rev. Biochem.* **53**, 293-321 (1984).
- Anderson, J. E., Ptashne, M. & Harrison, S. C. *Nature* **326**, 846-852 (1987).
- Aggarwal, A. K., Rodgers, D., Drott, M., Ptashne, M. & Harrison, S. C. *Science* (in the press).
- Ptashne, M. *A Genetic Switch* (Cell Press and Blackwell Scientific, Cambridge, Massachusetts, 1986).
- Ohlendorf, D. H., Anderson, W. F., Lewis, M., Pabo, C. O. & Matthews, B. W. *J. molec. Biol.* **169**, 757-769 (1983).
- Sauer, R. T., Yocum, R. R., Doolittle, R. F., Lewis, M. & Pabo, C. O. *Nature* **298**, 447-451 (1982).
- Mondragon, A., Subbiah, S., Almo, A., Drott, M. & Harrison, S. C. *J. molec. Biol.* (in the press).
- Wolberger, C. & Harrison, S. C. *J. molec. Biol.* **196**, 951-954 (1987).
- Durbin, R. M. *et al. Science* **232**, 1127-1137 (1986).
- Mondragon, A., Wolberger, C. & Harrison, S. C. *J. molec. Biol.* (in the press).
- Sussman, J. L., Holbrook, S. R., Church, G. M. & Kim, S. H. *Acta crystallogr.* **A33**, 800-804 (1977).
- Dickerson, R. E., Weinzierl, J. E. & Palmer, R. A. *Acta crystallogr.* **B24**, 997-1003.
- Bricogne, G. *Acta crystallogr.* **A132**, 832-847 (1976).

- Rossmann, M. G. & Blow, D. M. *Acta crystallogr.* **16**, 39-45 (1973).
- Bricogne, G. *Acta crystallogr.* **A130**, 395-405 (1974).
- Gaykema, W. P. J. *et al. Nature* **309**, 23-29 (1984).
- Rossmann, M. G. *et al. Nature* **317**, 145-153 (1985).
- Hogle, J. M., Chow, M. & Filman, D. J. *Science* **229**, 1358-1365 (1985).
- Rayment, I., Baker, T. S. & Casper, D. L. D. *Acta crystallogr.* **B39**, 505-514 (1983).
- Luo, M. *et al. Science* **235**, 182-191 (1987).
- Jones, T. A. *J. appl. Crystallogr.* **11**, 268-272 (1978).
- Sheriff, S. & Hendrickson, W. A. *Acta crystallogr.* **A43**, 118-1121 (1987).
- Anderson, J. E., Ptashne, M. & Harrison, S. C. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1307-1311 (1984).
- Koudelka, G. B., Harbury, P., Harrison, S. C. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Bushman, F. D. thesis, Harvard Univ. (1988).
- Bushman, F. D., Anderson, J. E., Harrison, S. C. & Ptashne, M. *Nature* **316**, 651-653 (1985).
- Wharton, R. P., Brown, E. L. & Ptashne, M. *Cell* **38**, 361-369 (1984).
- Koudelka, G. B., Harrison, S. C. & Ptashne, M. *Nature* **326**, 886-888 (1987).
- Wharton, R. P. & Ptashne, M. *Nature* **326**, 888-891 (1987).
- Crawford, J. thesis, Harvard Univ. (1977).
- Winkler, F. K., Schutt, C. E. & Harrison, S. C. *Acta crystallogr.* **A35**, 901-911 (1979).

The T-cell repertoire is heavily influenced by tolerance to polymorphic self-antigens

Ann M. Pullen*, Philippa Marrack*[†] & John W. Kappler*[§]

* Howard Hughes Medical Institute and Division of Basic Immunology, Department of Medicine, National Jewish Center for Immunology and Respiratory Medicine, 1400 Jackson Street, Denver, Colorado 80206, USA

[†] Departments of Biochemistry, Biophysics and Genetics, [§] Departments of Microbiology, Immunology and Medicine, University of Colorado Health Sciences Center, Denver, Colorado 80206, USA

T cells with $V_{\beta}3^{+}\alpha\beta$ receptors are deleted by self-tolerance in mice with particular major histocompatibility complex/self-antigen combinations. This also occurs for other V_{β} elements. Polymorphisms in the major histocompatibility complex and/or the self-antigens that cause massive deletion of T cells using particular V_{β} elements may be maintained by the need to balance the advantage of a diverse T-cell repertoire against the potential involvement of those elements in autoimmune disease.

DURING an immune response to a foreign antigen, T cells whose $\alpha\beta$ receptors recognize fragments of the antigen complexed to major histocompatibility complex (MHC) products are stimulated¹⁻³. The specificity of receptor binding is determined by residues in the antigen and MHC portions of the ligand. Polymorphic residues in the MHC molecule are particularly important both in complexing antigen and binding to the receptor³⁻⁶. The repertoire of $\alpha\beta$ receptors carried by mature T cells is shaped by the interaction of developing T cells with self-antigens and the products of particular MHC alleles expressed in that host. This shaping of the repertoire is thought to occur in at least two ways. First, in the thymus, immature T cells appear whose $\alpha\beta$ receptors are apparently random combinations of the germline-encoded variable elements (V_{β} , D_{β} , J_{β} , V_{α} , J_{α})⁷⁻¹⁰. Those T cells whose receptors are capable of interacting with the self-MHC expressed on thymic epithelial cells undergo positive selection for further differentiation^{11,12}. The mechanism of this selection is poorly understood, but it results in the enrichment of T cells with receptors capable of binding antigen/self-MHC complexes. Secondly, in order to achieve a repertoire tolerant of the host antigens, T cells that are reactive to self-antigens complexed to self-MHC are deleted from this pool^{8,13-15}.

The importance of positive selection in shaping the T-cell repertoire has been documented in various ways^{11,12,16}, whereas

the extent to which clonal deletion limits the repertoire during induction of self-tolerance has been difficult to measure, because of both the vast heterogeneity of $\alpha\beta$ receptors and the complexity involved in the interaction of the receptor variable elements with the antigen/MHC ligand. Recently we observed a simplified situation in the mouse in which virtually all $\alpha\beta$ receptors which used the variable element $V_{\beta}17a$ recognized a ligand composed of a B-cell-encoded antigen(s) in combination with the class II MHC molecule, I-E (refs 8 and 17). This reactivity led to almost complete depletion of $V_{\beta}17a^{+}$ T cells by self-tolerance in strains of mice expressing I-E, whereas in I-E⁻ strains 5-15% of T cells were $V_{\beta}17a^{+}$. Thus in this case, tolerance to a self-antigen/MHC ligand had an obvious and dramatic influence on the T-cell repertoire.

We assumed at first that $V_{\beta}17a$ would turn out to be a special case; but now we and others have examined more V_{β} elements, and the elimination of T cells bearing a particular V_{β} element by self-tolerance to a particular polymorphic self-antigen/MHC ligand seems to be the rule rather than the exception^{13,14,18}. Here, we provide evidence for this in the case of $V_{\beta}3$. Expression of $V_{\beta}3$ varies widely in mice, and is controlled mainly by tolerance to the combination of polymorphic self-antigen(s) and particular alleles of the murine MHC, *H-2*. These results not only establish the importance of tolerance to self-antigens in shaping the T-cell repertoire, but also pose the question of why alleles of those self-antigens that can eliminate such a high proportion of the T-cell repertoire are maintained in the population during evolution.

Production of an anti- $V_{\beta}3$ antibody

We purified $V_{\beta}3$ -containing $\alpha\beta$ receptor from lysates of a $V_{\beta}3^{+}$ T-cell hybridoma, 2B4.6 (ref. 19), using an immunoaffinity column bearing a monoclonal antibody (A2B4) specific for the α -chain of this receptor²⁰. Two Armenian hamsters were immunized many times with the purified receptor. Sera from both animals immunoprecipitated the $\alpha\beta$ receptor from lysates of 2B4.6, but not from a $V_{\beta}3^{-}$ T-cell hybridoma, DO-11.10 (data not shown). B-cell hybrids were produced from one of the animals by fusion of spleen cells to the variant myeloma line P3X63AG8.653. The primary hybrid supernatants were screened for binding to 2B4.6, but not to the parental tumour line, BW5147. KJ25a, one of three monoclonal antibodies with this specificity, was cloned and used in further studies. To determine the specificity of KJ25a, we examined a panel of T-cell hybridomas from a variety of strains and with different specificities, with known V_{α} and V_{β} usage (Table 1). Using flow cytometric analysis, we established that KJ25a binding correlated with $V_{\beta}3$ expression and also that this binding occurred regardless of the α -chain expressed by the T-cell hybrids.

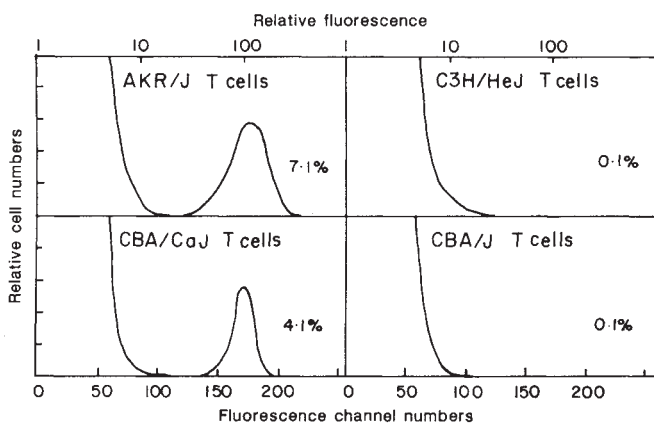


Fig. 1 Binding of KJ25 to lymph node T cells. Nylon-fibre-purified T cells⁴³ from the strains indicated were stained using biotinylated KJ25 and phycoerythrin streptavidin (Becton Dickinson) as in Table 1. 50,000 cells were analysed for each strain. The percentage of cells staining brighter than the controls (no primary antibody) is shown.

Table 1 KJ25a binds $V_{\beta}3^{+}$ $\alpha\beta$ receptors regardless of the V_{α} present

Hybrid	Strain	Specificity	V_{α}	V_{β}	Relative fluorescence	
					KJ25a	anti-T3
2B4.6	B10.A	pigeon cytochrome <i>c</i>	11	3	1.0	1.0
4HB72	C57BL/10	unknown	2	3	0.9	1.9
1HB117	C57BL/10	unknown	3	3	1.1	1.6
18BBM11	C57BL/10	I-A ^{bm12}	4	3	1.3	1.0
4BR20	B10.BR	unknown	8	3	1.5	1.4
1HB20	C57BL/10	unknown	DO	3	0.9	3.0
2BR33	B10.BR	unknown	11	8	0.0	1.8
2B23-23	C57L	unknown	11	17a	0.0	2.9
3DT-52.5	BALB/c	H-2D ^d	1	8.1	0.1	0.7
2B23-13	C57L	unknown	8	17a	0.0	2.2
DO-11.10	BALB/c	cOVA/I-A ^d	DO	8.2	0.0	2.9

T-cell receptor α - and β -chain usage of T-cell hybridomas was assigned from complementary DNA sequences (2B4.6, 3DT-52.5, DO-11.10, α and β), by reactivity of the hybridomas with anti- $V_{\beta}8$ or anti- $V_{\beta}17a$ monoclonal antibodies^{8,28}, or by dot blot analysis of hybridoma RNA with a collection of probes specific for V_{α} and V_{β} elements. In the latter case hybrids were chosen which hybridized with only one V_{α} and V_{β} probe (in addition to the $V_{\alpha}1/V_{\beta}1$ hybridizations to the transcripts derived from the parent BW5147 tumour line¹⁸). Reactivity with KJ25a or with the anti-CD3 antibody⁴⁴, 145-2C11, (kindly provided by Dr J. Bluestone) was determined with biotinylated antibody and phycoerythrin-coupled streptavidin (Becton-Dickinson)⁸. Analysis was using an Epics C flow cytometer as previously described⁸. Mean channel fluorescence is expressed relative to the $V_{\beta}3^{+}$ T-cell hybridoma, 2B4.6, a generous gift of Dr S. Hedrick, which we recloned, selecting for high receptor expression.

Control of $V_{\beta}3$ expression

In our previous experiments with $V_{\beta}17a$ and $V_{\beta}8.1$ and those of others with $V_{\beta}6$, the correlation between T-cell receptors containing these elements and reactivity to a particular self-antigen/H-2 combination was established by the clonal deletion of T cells bearing these V_{β} elements in mice carrying the appropriate genes^{8,13,14}. To test whether $V_{\beta}3$ would exhibit similar properties, we used flow cytometric analysis to examine the levels of $V_{\beta}3^{+}$ T-cells in a wide variety of strains of different H-2 types and genetic backgrounds. Figure 1 shows sample fluorescent histograms for the KJ25a-staining of purified lymph node T cells from four strains and the cumulative data are shown in Table 2.

We found a large variation in the level of $V_{\beta}3^{+}$ T-cells in these strains. For example, AKR/J and CBA/Ca mice had substantial numbers of $V_{\beta}3^{+}$ T cells, but C3H/HeJ and CBA/J had virtually none (Fig. 1). The summarized data in Table 2 show that the level of $V_{\beta}3^{+}$ T cells was controlled both by H-2 and non H-2-linked genes. For example among I^k mice, AKR/J, CBA/Ca, B10.BR and C57BR mice had high levels of $V_{\beta}3^{+}$ T cells, whereas C3H/HeJ, CBA/J, BALB.K and A/J had very few $V_{\beta}3^{+}$ T cells. Similarly among I^d mice B10.D2 contained $V_{\beta}3^{+}$ T cells, but DBA/2, BALB/c and NZB had deleted nearly all these cells. These results indicate that non-H-2-linked genes have a profound effect on $V_{\beta}3^{+}$ T-cell depletion suggesting that the presence of a polymorphic self-antigen(s) led to the deletion of $V_{\beta}3^{+}$ T cells.

The importance of H-2 products in the deletion was established with H-2 congenic mice. For example, when a set of H-2 congenic mice on the related D1 and D2 backgrounds was examined, there was a hierarchy of $V_{\beta}3^{+}$ expression depending on the H-2 type. Unlike DBA/2 (H-2^d) mice in which the deletion of $V_{\beta}3^{+}$ T cells was virtually complete, DBA/1(H-2^q) mice had a substantial number of $V_{\beta}3^{+}$ T cells, whereas both D1.LP (H-2^b) and D2.GD (H-2^{d/b}) mice had an intermediate level. Similar results were obtained with mice with the BALB or C3H background. There were an intermediate number of $V_{\beta}3^{+}$ T cells in H-2^b mice and virtually none in H-2^d or H-2^k mice. It is interesting that this hierarchy (H-2^k, H-2^d > H-2^b > H-2^q) is the same as that for the deletion of $V_{\beta}8.1^{+}$ T cells¹³ by the self-antigen, Mls-1^a (refs 13, 21), although the high level of $V_{\beta}3^{+}$ T cells in AKR/J (Mls-1^e) mice excludes Mls-1^a as an important locus in $V_{\beta}3^{+}$ T-cell deletion. Finally, as expected, H-2 had little effect on $V_{\beta}3$ expression in the H-2 congenic

series on the B10 background, because these mice seem to lack the structural gene(s) for the required non-H-2 self-antigen.

Thus, as with $V_{\beta}17a$, $V_{\beta}8.1$ and $V_{\beta}6$ (refs 8, 13, 14), mice carrying the appropriate alleles for H-2 and a particular non-H-2 self-antigen(s) eliminate virtually all T cells bearing $V_{\beta}3$. The

Table 2 Control of $V_{\beta}3$ expression by H-2 and Mls-2

Strain	H-2				Mls-1	Mls-2	% Peripheral T cells expressing $V_{\beta}3$
	K	A	E	D			
B10.BR	k	k	k	k	b	b	3.2±0.5
C57BR	k	k	k	k	b	b	7.2±0.4
AKR/J	k	k	k	k	a	b	6.7±0.2
CBA/CaJ	k	k	k	k	b	b	4.1±0.2
CBA/J	k	k	k	k	a	a	0.1±0.0
BALB.K	k	k	k	k	b	a	0.1±0.0
C3H/HeJ	k	k	k	k	b	a	0.1±0.0
A/J	k	k	k	d	b	a	0.1±0.0
B10.D2	d	d	d	d	b	b	4.0±0.8
DBA/2	d	d	d	d	a	a	0.1±0.0
BALB/c	d	d	d	d	b	a	0.1±0.0
NZB	d	d	d	d	a	a	0.1±0.0
C57BL/10	b	b	-	b	b	b	2.3±0.3
C57BL/6	b	b	-	b	b	b	3.3±0.1
C57L	b	b	-	b	b	b	6.1±0.2
D1.LP	b	b	-	b	a	a	1.0±0.2
BALB.B	b	b	-	b	b	a	1.7±0.4
C3H.SW	b	b	-	b	b	a	1.5±0.1
B10.Q	q	q	-	q	b	b	2.5±0.7
SWR	q	q	-	q	?	b	3.3±0.2
DBA/1	q	q	-	q	a	a	2.6±0.2
B10.PL	u	u	u	u	b	b	2.8±0.6
B10.M	f	f	-	f	b	b	2.5±0.4
B10.S(7R)	s	s	-	d	b	b	4.1±0.4
SJL	s	s	-	s	b	?	5.2±0.1
D2.GD	d	d	-	b	a	a	1.1±0.1

Purified T cells from the strains indicated were analysed by staining as described in Fig. 1. Three individual mice were analysed on different days and the results are expressed as mean ± standard error of the mean. All mice were bred in our own animal facility or were obtained from the Jackson Laboratory, Bar Harbour, Maine. Mls assignments were made according to Abromson-Leeman *et al.*²⁶, with slightly changed nomenclature as suggested²⁷.

Table 3 Non-*H-2* genes control $V_{\beta}3$ expression in AKXD recombinant inbred mice

Strain	<i>H-2</i>	<i>MTV-13</i>	<i>Mpmv-17</i>	% Peripheral T cells Expressing $V_{\beta}3$
DBA/2	d	d	d	0.1 ± 0.0
AKR/J	k	a	a	6.7 ± 0.2
AKXD-14	d	a	a	6.7 ± 0.4
AKXD-10	k	a	a	7.1 ± 0.3
AKXD-3	k	a	a	0.5 ± 0.1
AKXD-2	d	d	a	2.8 ± 0.6
AKXD-23	d	d	a	2.1 ± 0.3
AKXD-13	d	d	a	2.1 ± 0.3
AKXD-18	d	d	a	1.9 ± 0.2
AKXD-16	d	d	a	1.9 ± 0.2
AKXD-8	d	d	a	1.1 ± 0.2
AKXD-9	d	d	a	1.1 ± 0.2
AKXD-26	k	d	a	1.6 ± 0.1
AKXD-15	k	d	a	1.2 ± 0.2
AKXD-1	k	d	a	0.8 ± 0.1
AKXD-25	k	d	a	0.4 ± 0.1
AKXD-11	k	d	a	0.3 ± 0.1
AKXD-27	k	d	a	0.3 ± 0.1
AKXD-24	d	a	d	0.4 ± 0.1
AKXD-20	d	d	d	0.1 ± 0.1
AKXD-21	k	a	d	0.1 ± 0.1
AKXD-6	k	d	d	0.1 ± 0.1

Purified T cells from the AKXD recombinant inbred strains indicated were stained as described for Fig. 1. Three individual mice were analysed on different days and the results are expressed as mean ± standard error of the mean. The mice were obtained from the Jackson Laboratory, Bar Harbour, Maine. *MTV-13* is a mouse mammary tumour virus polymorphism detected as a restriction fragment length polymorphism with a long terminal repeat probe and *EcoRI*. It has been mapped to chromosome 4 (Drs J. Angel and B. A. Taylor, personal communication). *Mpmv-17* is an endogenous modified polytropic virus detected by Southern blot analysis and carried by AKR, but not DBA/2. Its chromosomal location is unknown, but it has been mapped to a linkage group including the gene for the lymphocyte *Ly7* antigens (W. N. Frankel, J. P. Stoye, B. A. Taylor and J. M. Coffin, manuscript in preparation).

strain distribution of the self-antigen(s) controlling depletion of $V_{\beta}3^{+}$ has a striking resemblance to the *Mls-2^a* gene(s) (formerly *Mls^c*)²¹⁻²³. Originally thought to be an allele of *Mls-1*, the product of this gene(s) was defined in C3H (*H-2^k*) mice by the ability of spleen cells from this strain to stimulate T cells from other *H-2^k* mice in a primary mixed lymphocyte reaction. Subsequent experiments indicated that this gene(s) is not linked to *Mls-1* on chromosome 1 and the name *Mls-2* has been proposed²⁴⁻²⁷. The genetics of *Mls-2* are still incomplete and its definition has not been standardized. Several laboratories have attempted to define the locus by the response of anti-C3H T-cell clones to spleen cells from various parental and recombinant strains. The strain distribution and *H-2* dependence of the antigen stimulating some of these clones is identical to that causing the depletion of $V_{\beta}3^{+}$ T cells^{24,25}.

We have begun to examine the genetics of this self-antigen(s) by analysing a set of recombinant mice inbred between AKR/J (*H-2^k*) and DBA/2 (*H-2^d*) (Table 3). Because the *H-2* types of these strains are permissive for $V_{\beta}3^{+}$ T-cell deletion, but only DBA/2 mice carry the gene(s) for the non-*H-2* antigen, the frequency of $V_{\beta}3^{+}$ T cells in these recombinant strains should indicate the genetics of *Mls-2*. The mice did not fall into two categories. $V_{\beta}3$ levels in a few of the recombinant strains were high, like AKR (~7%), or very low, like DBA/2 (0.1%); most had intermediate levels ranging only from 0.3% to 2.8%. Although these data are somewhat limited, the frequency of intermediates indicates that at least two, and probably more,

non-*H-2* genes contribute to the self-antigens deleting $V_{\beta}3^{+}$ T cells. The complexity of the genetics involved indicate that it is perhaps premature to accept a simple *Mls-2* designation for these genes.

With such a wide range of expression of $V_{\beta}3$ in these mice, it is difficult to deduce anything definitive about the complex genetics of *Mls-2*. But if the intermediate mice are grouped together, examination of the genes that have been studied in these recombinant inbred mice identifies two genes that may be linked to the deleting self-antigens. The first is *Mpmv-17*, an endogenous modified polytropic virus carried by AKR, but not DBA/2 mice. It is in a linkage group containing the *Ly7* gene. With one exception (AKXD-24), recombinants lacking this marker tend to delete $V_{\beta}3^{+}$ T cells completely, whereas the others have high or intermediate levels of $V_{\beta}3^{+}$ T cells. The second locus, *MTV-13*, is the mouse mammary tumour virus on chromosome 4. With the exception of AKXD-3, mice carrying the *Mpmv-17* marker have high levels of $V_{\beta}3^{+}$ T cells if they also have the AKR allele at *MTV-13* and intermediate levels if they have the DBA/2-derived *MTV-13* allele. Without more data, these results are inconclusive.

Assuming that the moiety responsible for clonal deletion of the self-reactive $V_{\beta}3^{+}$ T cells should also stimulate peripheral $V_{\beta}3^{+}$ T cells from strains which did not express this moiety, we hypothesized that a random set of $V_{\beta}3^{+}$ T-cell hybrids generated from AKR/J (*Mls-2^b*) mice would be stimulated at high frequency by spleen cells from the *Mls-2^a* strains which had deleted $V_{\beta}3^{+}$ T cells. We generated random KJ25a⁺ T-cell hybrids from AKR/J mice as previously described^{8,28} and tested 14 for reactivity to spleen stimulator cells from a variety of strains (Table 4). Most of the hybrids tested responded by interleukin-2 (IL-2) production to stimulation by C3H/HeJ, CBA/J or DBA/2 spleen cells, although the response with DBA/2 was weaker. As expected, cells from AKR or CBA/Ca mice did not stimulate the hybrids. These results confirm the presence in mice which have depleted $V_{\beta}3^{+}$ T cells of an antigen that can stimulate $V_{\beta}3^{+}$ T cells from other strains. The deletion of $V_{\beta}3^{+}$ T cells in these strains is virtually complete, but not all hybridomas responded to the appropriate spleen cells. This suggests either that self-tolerance is more sensitive than hybridoma stimulation, or that not all the self-antigens involved are well represented on normal spleen cells.

Transfer of self-antigens

The definition of the products of genes like *Mls-1* and *Mls-2* as antigens has been controversial. It has been suggested that these gene products do not encode or control the expression of unique self-antigens, but rather alter the function of the antigen-presenting cells in which they are expressed, such that low-affinity interactions between MHC and $\alpha\beta$ receptors are promoted^{22,29}. Because polymorphisms in both these genes and in *H-2* control the deletion of $V_{\beta}3$, $V_{\beta}6$ and $V_{\beta}8.1$ -bearing T cells^{8,13,14}, we were able to design an experiment to test this hypothesis. If in fact the non-*H-2* genes involved in these deletions control a property of the antigen-presenting cells, they would exert their effect only in those cells in which they were expressed. On the other hand, if these genes control expression of self-antigens, their product might be produced in one cell but picked up, processed and presented by other cells. Such a situation has been reported for minor histocompatibility antigens recognized by cytotoxic T cells or in graft versus host disease^{30,31} and, in one report, for *Mls-1^a* recognition³².

We constructed bone marrow chimaeric mice in which a mixture of bone marrow cells from DBA/1 (*H-2^q*) and B10.BR (*H-2^k*) were used to repopulate irradiated (B10.Q_βBR × B10.BR_βBR)_{F1} animals (*H-2^q* × *H-2^k*). In this case the functional *H-2^k* genes were expressed only in the B10.BR and *F1* cells, and the *Mls-1^a* and *Mls-2^a* genes required for V_{β} depletion were carried only by the DBA/1 (*H-2^q*) cells. Because *H-2^q* encoded products are particularly ineffective in the dele-

Table 4 $V_{\beta}3^{+}$ T-cell hybridomas from AKR/J often react with *Mls-2*^a

Hybrid	Relative fluorescence with KJ25a	IL-2 production (units ml ⁻¹) in response to			
		AKR/J (b)	CBA/CaJ (b)	C3H/HeJ (a)	DBA/2 (a)
K25-56	1.8	≤5	≤5	158	10
K25-20	1.5	≤5	≤5	124	13
K25-49	2.9	≤5	≤5	99	36
K25-59	3.3	≤5	≤5	78	7
K25-29	3.0	≤5	≤5	78	8
K25-28	2.9	≤5	≤5	61	8
K25-16	3.2	≤5	≤5	53	7
K25-34	2.3	≤5	≤5	24	≤5
K25-48	4.0	≤5	≤5	7	≤5
K25-57	3.1	≤5	≤5	7	≤5
K25-3	2.4	≤5	≤5	≤5	≤5
K25-8	1.5	≤5	≤5	≤5	≤5
K25-44	2.9	≤5	≤5	≤5	≤5
K25-45	2.6	≤5	≤5	≤5	≤5

A random panel of KJ25a⁺ T-cell hybridomas generated from AKR/J mice was stained with KJ25a as in Table 1. Fluorescent intensity is expressed relative to 2B4.6. The hybridomas were tested as previously described⁴⁵ for their ability to respond to spleen cells from various mouse strains by interleukin 2 (IL-2) production. *Mls-2* designations in parentheses as in Table 2. The results shown are the geometric average of one to three determinations.

tion of T cells with these V_{β} elements^{13,33}, deletion would require the acquisition of the DBA/1 *Mls* products by the H-2^k antigen-presenting cells. As controls we prepared chimaeric mice reconstituted with either DBA/1 or B10.BR bone marrow alone. Six to eight weeks after reconstitution the parental and chimaeric mice were examined for $V_{\beta}3$, $V_{\beta}6$ and $V_{\beta}8.1$ T cells and the extent of chimaerism in T cells and non-T cells established (Table 5).

Similar results were obtained for each V_{β} element. For example, with $V_{\beta}3$, there was good expression, as expected, in B10.BR, DBA/1 and the F₁ recipient strain. Likewise, in the B10.BR → F₁ mice a substantial number of $V_{\beta}3^{+}$ T cells were seen. But in the DBA/1 → F₁ and DBA/1 & B10.BR → F₁ mice, $V_{\beta}3^{+}$ T cells were severely depleted, indicating the acquisition of the appropriate gene products from DBA/1 cells by H-2^k-expressing B10.BR and F₁ cells. The extent of the deletion of $V_{\beta}3^{+}$ T cells in DBA/1 → F₁ mice was noteworthy, as we could detect few or no class II MHC-expressing H-2^k cells in the lymph nodes of these mice. The results indicate either that the process of deletion requires only a few H-2^{k+} cells or that F₁ recipient cells other than those derived from bone marrow can participate in the process. Reciprocally, one of the DBA/1 & B10.BR → F₁ mice was very poorly reconstituted with DBA/1 bone marrow at the time of death. Nevertheless, deletion of $V_{\beta}3^{+}$ T cells was just as complete, again suggesting that the process of deletion is very efficient. These results are similar to those we reported previously with $V_{\beta}17a$ (ref. 34), in which transgenic mice carrying a MHC *Eα* gene construct which allowed for I-E expression on only a small fraction of B cells, nevertheless showed maximal deletion of $V_{\beta}17a^{+}$ T cells.

The results for $V_{\beta}6$ and $V_{\beta}8.1$ were similar. Because the F₁ host was constructed to lack the structural genes for the $V_{\beta}8$ family³⁵, all $V_{\beta}8.1^{+}$ T cells in the chimaeras had to arise from the donor bone marrow. The results, however, were the same in either case. The presence of H-2^k cells and *Mls-1*^a DBA/1 cells in the same host resulted in the very efficient deletion of both $V_{\beta}6^{+}$ and $V_{\beta}8.1^{+}$ T cells, although the depletion may have been more complete in the double bone marrow chimaera. This may be because dilution of H-2^k gene products on F₁ cells led to somewhat less effective expression of *Mls-1*^a plus *I*^k in DBA/1 → F₁ chimaeras than in DBA/1 & B10.BR → F₁ mice. As a control, we determined $V_{\beta}8.2$ and $V_{\beta}8.3$ expression and found, as expected, little difference among the mice. These results clearly demonstrate the passive acquisition of the functional

gene products of both *Mls-1*^a and *Mls-2*^b and provide good evidence that these genes do not control an innate property of the antigen-presenting cells in which they are expressed. Although they do not formally show that these genes encode peptide antigens, this is by far the most straightforward conclusion.

Discussion

The reactivity of nearly all $V_{\beta}17a^{+}$ T cells towards a product(s) of B cells seen in association with the I-E molecule provided a useful tool in the study of self-tolerance^{8,17}. It seemed at the time that this correlation between a V_{β} element and self-reactivity would be unusual. But of the six, out of about twenty, murine V_{β} elements that have now been examined in some detail ($V_{\beta}3$, $V_{\beta}6$, $V_{\beta}8.1$, $V_{\beta}8.2$, $V_{\beta}8.3$ and $V_{\beta}17a$), all but $V_{\beta}8.2$ and $V_{\beta}8.3$ are deleted in strains of mice carrying particular *H-2* alleles and self-antigens^{8,13,14}. In addition, there are indications that other V_{β} elements (for example $V_{\beta}5$, $V_{\beta}11$, $V_{\beta}12$) will be associated with similar phenomena¹⁸.

What is the significance of the unexpected result that for many, if not most, of the V_{β} elements, self-antigens exist in the species that will cause the deletion of most T cells expressing these elements regardless of the other variable components in the receptor? In the many responses to foreign antigens which have been examined, the repertoire of $\alpha\beta$ receptors involved is generally very complex, using many V_{β} elements³⁶. Even in those cases where there are well defined antigens which raise a response in which use of a particular V_{β} element predominates, the frequency of responding clones is generally very low, and the specificity is highly dependent on the other variable elements in the receptor^{37,38}. The answer to this paradox may lie in the extensive diversity of self-antigens to which self-tolerance must be established. It is perhaps not surprising that among this huge array of potential antigens, there are a few which, when complexed with MHC molecules, fortuitously display an array of amino-acid residues that interact so strongly with the V_{β} portion of the receptor that the requirement for interactions with the other portions of the receptor is minimal. Because the tolerance-inducing effect of such self-antigens would be so dramatic, we are bound to observe them if they exist. Such reasoning might predict that there would be a similar phenomenon for V_{α} elements. This possibility has been hard to examine however, because of the difficulty in producing antibodies to V_{α} elements and because V_{α} genes exist in related, cross-hybridizing gene families.

Table 5 *In vivo* transfer of self-antigens

	Mouse	% Chimaerism of lymph node cells						% T cells expressing V_{β}				
		Non-T cells			T cells							
		DBA/1	B10.BR	F_1	DBA/1	B10.BR	F_1	3 (% Total T cells)	6 (% Donor T cells)	8.1	8.2	8.3
DBA/1	1							2.6	3.7	5.1	11.7	8.8
	2	100	—	—	100	—	—	3.2	5.1	5.1	12.2	6.8
	3							2.8	5.1	5.2	11.9	9.6
B10.BR	1							3.2	10.1	5.6	12.8	8.5
	2	—	100	—	—	100	—	3.9	11.3	7.5	12.3	7.6
	3							4.4	11.5	6.9	13.2	7.7
(B10.Q $_{\beta\text{BR}}$ × B10.BR $_{\beta\text{BR}}$) F_1	1							4.2	15.9	—	—	—
	2	—	—	100	—	—	100	3.5	16.5	—	—	—
	3							3.9	16.8	—	—	—
B10.BR → F_1	1	—	>95	<5	—	58	42	5.5	10.8	6.1	14.4	8.2
	2	—	>95	<5	—	60	40	4.5	11.2	6.3	15.7	11.0
	3	—	>95	<5	—	53	47	4.5	11.9	6.3	15.6	9.0
DBA/1 → F_1	1	>95	—	<5	71	—	29	0.2	1.5	3.1	14.0	12.3
	2	>95	—	<5	68	—	32	0.3	1.6	2.4	15.9	14.9
	3	93	—	7	73	—	27	0.3	2.0	3.0	15.6	12.5
DBA/1 and B10.BR → F_1	1	<5	>95	<5	2	71	27	0.2	0.6	1.6	16.7	12.1
	2	80	10	10	68	16	16	0.2	1.1	1.4	17.9	15.1
	3	85	15	<5	63	4	33	0.2	0.9	1.9	17.3	15.9

F_2 mice were produced between B10.Q (H-2^q) which carry the V_{β} haplotype with about 20 V_{β} elements and C57BR (H-2^k) mice which have a related genetic background, but carry the V_{β} haplotype in which a large deletion has eliminated about half of the V_{β} elements including the $V_{\beta}8$ family³⁵. The F_2 mice which were homozygous H-2^q and homozygous V_{β} were used to establish a strain B10.Q $_{\beta\text{BR}}$. Similarly, a strain was established from B10.BR (H-2^k) and C57BR mice that was homozygous H-2^k and V_{β} , B10.BR $_{\beta\text{BR}}$. These and other V_{β} homozygous strains will be described in more detail elsewhere (J.W.K., E. Kushnir and P.M., manuscript in preparation). Bone marrow chimaeric mice were established as previously described⁴⁶. Irradiated recipient (B10.Q $_{\beta\text{BR}}$ × B10.BR $_{\beta\text{BR}}$) F_1 mice were reconstituted with either DBA/1(H-2^q) bone marrow, B10.BR(H-2^k) bone marrow or a mixture of both. Six to eight weeks after reconstitution lymph node cells were prepared from the chimaeric mice and from the normal donor and recipient strains. T cells were prepared on nylon-fibre columns⁴³. The percentage of T cells derived from donor and recipient were determined by staining with appropriate monoclonal anti-class I H-2 antibodies. In addition the percentage of non-T cells derived from donor and recipient were estimated using two-colour staining with either anti-Thyl or anti-immunoglobulin M against appropriate anti-class-I or class II monoclonal antibodies. T cells were analysed for the percentage $V_{\beta}3$ expression using KJ25. In addition $V_{\beta}6$ was analysed using the $V_{\beta}6$ -specific monoclonal antibody RR-4.7 (O. Kanagawa, E. Palmer and J. Bill, submitted). The results are presented as the percentage of total T cells expressing these elements. $V_{\beta}8.1$, $V_{\beta}8.2$ and $V_{\beta}8.3$ expression was calculated as previously described¹³ using the monoclonal antibodies, KJ16, F23.1 and F23.2. In this case the results are reported as the percentage of donor T cells expressing these V_{β} elements because the F_1 (V_{β}^a) recipient lacked the structural genes for the $V_{\beta}8$ family. Three mice of each type were analysed, each on a different day.

These results with V_{β} -deleting self-antigens offer insight into the nature of the Mls antigens. The phenomena associated with Mls have puzzled immunologists for years. It was originally defined in the mouse as a set of non-H-2 alloantigens which, unlike other antigens, were capable of stimulating a very large proportion of MHC-identical T cells^{21,23}. The best studied example of an Mls antigen is the product of the *a* allele of the *Mls-1* locus on chromosome 1. What was previously thought to be the *c* allele of this locus is in fact coded elsewhere and has recently been referred to as *Mls-2^a* (refs 21–27). Our results and those of others¹⁴ now indicate that the Mls antigens are in fact examples of these tolerance-inducing self-antigens, identified because of their polymorphic nature, which when paired with certain MHC molecules react with an entire set of T cells expressing particular V_{β} elements ($V_{\beta}8.1$ and $V_{\beta}6$ for *Mls-1^a* and $V_{\beta}3$ for *Mls-2^a*). Using the sensitive approach of T-cell deletion, rather than T-cell stimulation, it is probable that additional *Mls*-like loci will be identified.

Why do polymorphic forms of MHC and self-antigens with the capacity to delete such large portions of the $\alpha\beta$ repertoire not succumb to evolutionary pressure, as presumably mice with a larger and more diverse T-cell receptor repertoire would have a distinct advantage in the wild? This question also applies to the continued maintenance in the mouse population of chromosomes carrying deletions or inactivations of genes encoding V_{β} or J_{β} elements^{35,39,40}. We suggest that it could sometimes be

advantageous for some mice in the population to lack certain types of T-cell receptors.

Current evidence indicates that induced autoimmunity can be driven by T cells with quite restricted repertoires. For example, experimental allergic encephalitis, a model for a multiple sclerosis-like disease inducible in H-2^u mice, may be caused chiefly by T cells expressing $V_{\beta}8.2$ (ref. 41). Induced rheumatoid arthritis in H-2^q animals may be caused by T cells bearing $V_{\beta}6$ (C. David, personal communication). Under some circumstances, the processes of normal tolerance induction by the target tissue seem unable to delete these potentially damaging self-reactive clones of T cells. But presumably mice resistant to the disease still exist in the population because the offending V_{β} element is absent, either through gene deletion or inactivation, or through T-cell elimination by tolerance to one of the V_{β} reactive self-antigens. In the population as a whole, these various stable polymorphisms in V_{β} , MHC and other self-antigens may balance the need for a diverse T-cell repertoire against the need to avoid autoimmunity. A theory of this type may account for the fact that certain human *HLA* alleles provide protection against the onset of juvenile diabetes⁴².

We thank Dr Edward Palmer for discussion and for the probes used in Northern blot analyses, Dr Osami Kanagawa for the anti- $V_{\beta}6$ antibody (RR-4.7), Dr Michael Bevan for the anti- $V_{\beta}8$ and anti- $V_{\beta}8.2$ antibodies (F23.1 and F23.2) and Dr Steven Hedrick for the 2B4 and A2B4 cell lines. Thanks also to Marcy

Blackman, Jerome Bill and Virginia Appel for many of the $\alpha\beta$ -characterized T-cell hybridomas in Table 1 and to Drs Angel, Frankel, Stoye, Coffin and Taylor for making their data on the AKXD recombinant inbred mice available before publication.

Received 23 June; accepted 13 September 1988.

1. Marrack, P. & Kappler, J. *Science* **238**, 1073-1079 (1987).
2. Babbitt, B., Allen, P., Matsueda, G., Haber, E. & Unanue, E. *Nature* **317**, 359-361 (1985).
3. Buus, S., Sette, A., Colon, S., Miles, C. & Grey, H. *Science* **235**, 1353-1358 (1987).
4. Bjorkman, P. J. *et al. Nature* **329**, 506-512 (1987).
5. Bjorkman, P. J. *et al. Nature* **329**, 512-518 (1987).
6. Ajitkumar, P. *et al. Cell* (in the press).
7. Roehm, N. *et al. Cell* **38**, 577-584 (1984).
8. Kappler, J., Roehm, N. & Marrack, P. *Cell* **49**, 273-280 (1987).
9. Snodgrass, R., Dembic, Z., Steinmetz, M. & von Boehmer, H. *Nature* **315**, 232-233 (1985).
10. Arden, B., Klotz, J., Siu, G. & Hood, L. *Nature* **316**, 783-787 (1985).
11. Bevan, M. & Fink, P. *Immunol. Rev.* **42**, 3-19 (1978).
12. Zinkernagel, R. M. *et al. J. exp. Med.* **147**, 882-896 (1976).
13. Kappler, J., Staerz, U., White, J. & Marrack, P. *Nature* **332**, 35-40 (1988).
14. MacDonald, H. R. *et al. Nature* **332**, 40-44 (1988).
15. Lo, D., Ron, Y. & Sprent, J. *Surg. Immun. Res.* **5**, 221-232 (1986).
16. Sprent, J. *J. exp. Med.* **147**, 1159-1174 (1978).
17. Marrack, P. & Kappler, J. *Nature* **332**, 840-842 (1988).
18. Bill, J., Apple, V. & Palmer, E. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
19. Hedrick, S., Cohen, D., Nielson, E. & Davis, M. M. *Nature* **308**, 149-153 (1984).
20. Samelson, L., Germain, R. & Schwartz, R. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6972-6976 (1983).
21. Festenstein, H. *Transplant. Rev.* **15**, 62-88 (1973).
22. Katz, M. E. & Janeway, C. A. *J. Immun.* **134**, 2064-2070 (1985).
23. Festenstein, H., Bishop, C. & Taylor, B. A. *Immunogenetics* **5**, 357-361 (1977).
24. Click, R. *et al. J. Immun.* **139**, 321-325 (1987).

We are grateful to Ella Kushnir, Terri Wade, Janice White and Rhonda Richards for excellent technical assistance. This work was supported in part by research grants from the NIH and the American Cancer Society.

25. Abe, R., Ryan, J. J. & Hodes, R. J. *J. exp. Med.* **165**, 1113-1129 (1987).
26. Abromson-Leeman, S. R., Laning, J. C. & Dorf, M. E. *J. Immun.* **140**, 1726-1731 (1988).
27. Janeway, C. A., Jr., Fischer-Lindahl, K. & Hammerling, U. *Immun. Today* **9**, 125-126 (1988).
28. Kappler, J. *et al. Cell* **49**, 263-271 (1987).
29. Webb, S. R. & Sprent, J. *Int. Rev. Immun.* **1**, 151-182 (1986).
30. Matzinger, P. & Bevan, M. J. *Cell. Immun.* **33**, 92-100 (1977).
31. Korngold, R. & Sprent, J. *J. exp. Med.* **151**, 314-327 (1980).
32. De Kruffy, R. H., Ju, S.-T., Laning, J., Cantor, H. & Dorf, M. E. *J. Immun.* **137**, 1109-1114 (1986).
33. Lynch, D. H., Gress, R. E., Needleman, B. W., Rosenberg, S. A. & Hodes, R. J. *J. Immun.* **134**, 2071-2078 (1985).
34. Van Ewijk, W. *et al. Cell* **53**, 357-370 (1988).
35. Behlke, M., Chou, H., Huppi, K. & Loh, D. *Proc. natn. Acad. Sci. U.S.A.* **83**, 767-771 (1986).
36. Garman, R., Ko, J.-L., Vulpe, C. & Raulet, D. *Proc. natn. Acad. Sci., U.S.A.* **83**, 3987-3991 (1986).
37. Fink, P. J. *et al. Nature* **321**, 219-226 (1986).
38. Barth, R. *et al. Nature* **316**, 517-523 (1985).
39. Wade, T., Bill, J., Marrack, P., Palmer, P. C. & Kappler, J. *J. Immun.* (in the press).
40. Kotzin, B. Barr, V. & Palmer E. *Science* **229**, 167-171 (1985).
41. Acha-Orbea, H. *et al. Cell* **54**, 263-273 (1988).
42. Todd, J. A. *et al. Science* **240**, 1003-1009 (1988).
43. Julius, M., Simpson, E. & Herzenberg, L. *Eur. J. Immun.* **3**, 645-649 (1973).
44. Leo, O., Foo, M., Sachs, D. H., Samelson, L. E. & Bluestone, L. *Proc. natn. Acad. Sci., U.S.A.* **84**, 1374-1378 (1987).
45. Kappler, J., Skidmore, B., White, J. & Marrack, P. *J. exp. Med.* **153**, 1198-1214 (1981).
46. Kappler, J. & Marrack, P. *J. exp. Med.* **148**, 1510-1522 (1978).

LETTERS TO NATURE

Discovery of a nebula around PSR1957+20

Shrinivas R. Kulkarni & J. Jeff Hester

Palomar Observatory, 105-24, California Institute of Technology, Pasadena, California 91125, USA

The eclipsing binary pulsar PSR1957+20 is presumed to have been spun up to its present rotation period of 1.6 ms by an earlier episode of accretion. We report here the discovery of an H α emission nebula around PSR1957+20, the first nebula to be found around a 'recycled', and thus ancient, pulsar. We suggest that the H α emission arises in shocks driven into the interstellar medium by a relativistic wind from the pulsar. The nebular spectrum shows similarities to fast shocks in young supernova remnants like Tycho¹; the absence of the usual optical forbidden lines is explained by models of the excitation of neutral atoms drifting into the post-shock region², but the large observed Balmer decrement ($I(\text{H}\alpha)/I(\text{H}\beta) \geq 12$) is more mysterious. Between 1% and 10% of the rotational luminosity is absorbed in the region outlined by the H α emission, with the bulk of the relativistic wind flowing away from the pulsar. This nebula adds to a growing number of pulsar wind nebulae, like the one around PSR1951+32 in CTB80 (ref. 3).

The principal accessible store of energy in a pulsar is its rotational energy, which is liberated at a rate $\dot{E}_R \approx I\omega\dot{\omega}$, where $I \approx 10^{45}$ g cm² is the moment of inertia of the neutron star and ω is its angular velocity. For a long time astronomers have sought some detectable effect which would identify the deposition of this energy into the interstellar medium (ISM)⁴. Numerous searches were conducted in several radio frequency bands but without success⁵.

By means of the Einstein satellite, Helfand and coworkers⁶ detected soft X-ray emission from five nearby pulsars which they interpreted as having arisen in a synchrotron nebula generated by the pulsar wind. Because of the faintness of the sources, the shapes and the spectra were not very well determined. Consequently, alternative interpretations, especially for the non-

extended X-ray sources, have been advanced which invoke thermal emission from the surface of the neutron star⁷⁻⁹. The synchrotron nebula interpretation is probably secure for the extended emission associated with PSR1055-52, which happens to be the brightest source in the above sample.

The ease with which the optical line emission from the shocks driven by the pulsar wind nebula (hereafter PWN) around PSR1951+32 ($\dot{E}_R = 3.7 \times 10^{36}$ erg s⁻¹) could be detected in the 6,563-Å H α line¹⁰ convinced us that narrow-band imaging is an ideal method to detect such nebulae, second only to studying the X-ray emission. Consequently we started a survey of optical observations of all northern pulsars with large $\dot{E}_R$.

The survey was done with a focal reducing camera and a charge-coupled detector (CCD) on the 60" telescope at Palomar Observatory. Images were obtained with a narrow-band H α filter (6,564 Å/15 Å) and an 'off-line' continuum filter (6,450 Å/104 Å). To avoid problems from chip defects and internal reflections, each field was observed at two or more different pointings. The images were subjected to the usual process of bias subtraction followed by flat fielding. Using a grid of stars, the 'on-line' and the 'off-line' images were registered and normalized to the same strength of the continuum and subtracted to yield the line emission images. The planetary nebula NGC7027 (refs 11, 12) was used for the absolute calibration.

PSR1957+20 was one of the prime targets; it has a period P of 1.6 ms (ref. 13) and a period derivative $\dot{P} \approx 3 \times 10^{-20}$ s s⁻¹ (J. H. Taylor, personal communication), implying a large rotational luminosity, $\dot{E}_R = 2.9 \times 10^{35}$ erg s⁻¹. PWN1957+20 was immediately apparent in the raw data obtained on the night of 15 July 1988. The sum of all the line emission images is shown in Fig. 1. Having thus detected the nebula we sought a signature in images in the 6,730-Å S II line but found none, despite an effective integration time of 3,600 s. Subsequently, we obtained several spectra of the nebula using the same imaging optics but with a slit in the focal plane and a grism in the collimated beam. The combined spectrum is shown in Fig. 2. With the intention of studying the detailed features, we obtained, on 19 July 1988, narrow-band H α images with the 4-Shooter camera¹⁴ at the Cassegrain focus of the Hale 5-m telescope. A deep image in an H α filter is shown in Fig. 3, with the pulsar candidate¹⁵⁻¹⁸ arrowed.

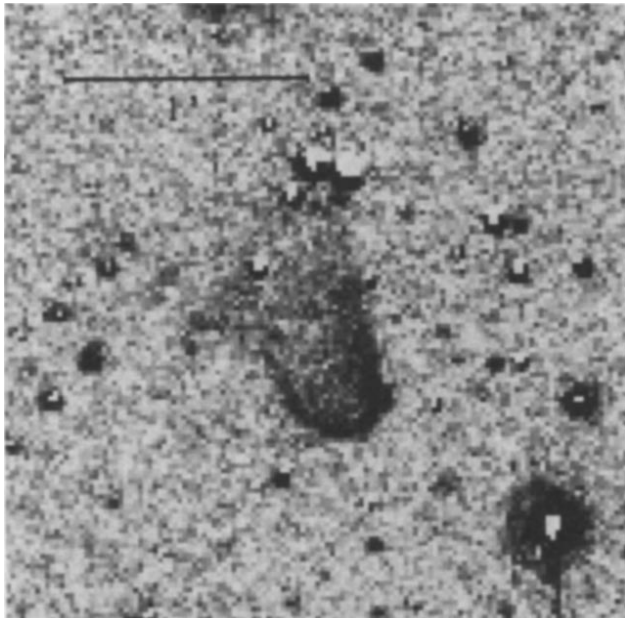


Fig. 1 A continuum-subtracted $H\alpha$ image of the field around PSR1957+20. North is up and east to the left. The data were obtained with a camera (effective f -ratio 1.65) at the Cassegrain focus of the Palomar 60" telescope followed by an 800² Texas Instrument CCD detector. Each pixel corresponds to a square of 1.2 arcsec side. The image shown here is a sum of three exposures representing a total of 5,400 s of integration and were obtained under conditions where the seeing was better than the resolution of the camera. Scale bar indicates 1 arcmin.

We suggest that the $H\alpha$ emission arises in the shock where the ISM reaches momentum balance with the relativistic pulsar wind. The distinctive cometary shape of the nebula indicates that ram pressure ($\rho_0 v_p^2$) (where ρ_0 is the density of the ambient medium and v_p is the pulsar velocity) dominates over the static ISM pressure $p_0 \approx 6 \times 10^{-13} n_0 T_0 / 4,000 \text{ erg cm}^{-3}$ (where n_0 is the ambient particle density (in cm^{-3}) and T_0 is the ambient temperature (in K)) and that the pulsar is moving approximately in the south-east direction with a speed that exceeds the sound speed in the ambient medium. Future timing observations should be able to verify the latter prediction. Indeed, interstellar scintillation observations indicate a transverse motion of 70 km s^{-1} (A. Wolszczan, personal communication); the uncertainty on such estimates is supposed to be better than a factor of two (ref. 19). We assume $v_p = 100 \text{ km s}^{-1}$, to take into account the unobserved radial component. Balancing the pulsar wind pressure ($\dot{E}_R / 4\pi R_w^2 c$) with the ram pressure, we find that R_w , the distance between the pulsar and the apex or the head of the nebula, is $0.03(n_0/\text{cm}^{-3})^{-1/2} \text{ pc}$. This value compares favourably with the

observed separation of 4 arcsec, equivalent to 0.02 pc if we adopt the dispersion-measure-derived distance of $\sim 1 \text{ kpc}$; projection effects make this agreement even better. Assuming a projection factor of ~ 0.5 yields $n_0 \approx 1 \text{ cm}^{-3}$.

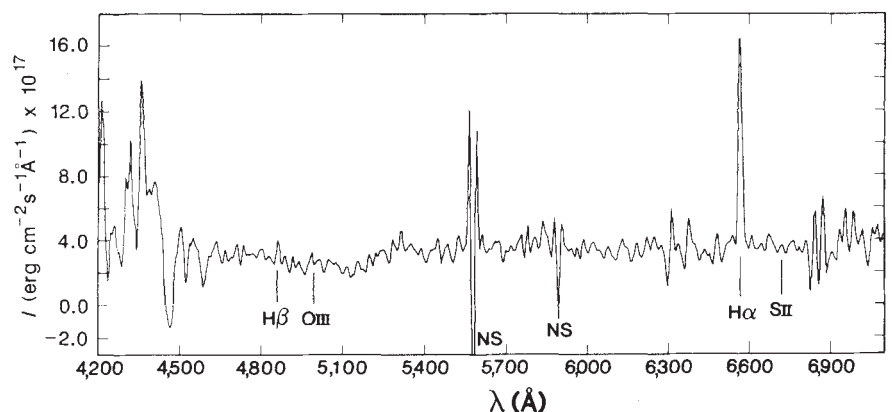
As expected, the mean pressure in the interior of the nebula, $p_n \approx 1.5 \times 10^{-10} n_0 \text{ erg cm}^{-3}$, is nearly three orders of magnitude higher than p_0 ; it is, however, comparable with that in PWN1055-52 (ref. 20). The high pressure is thought to be due to a relativistic, thermal plasma filling the nebula. In one model²¹, the pulsar wind is assumed to be thermalized at radii comparable with the light cylinder whereas in another model²² the pulsar wind is cold close to the pulsar but becomes thermalized at a distance $r \approx R_w$. Regardless of this uncertainty, we expect synchrotron emission from the interior of the bubble.

All evolutionary models²³⁻²⁵ for PSR1957+20 require it to be an old pulsar ($t > 10^9 \text{ yr}$), as is the case for all other millisecond pulsars²⁶. This is the first detection of a nebula around an ancient neutron star. Such a nebula was already anticipated²², however, following the discovery of the first 1.6-ms pulsar, 1937+214, and the predictions made then are applicable here. In particular, $\sim \frac{1}{2} \dot{E}_R$ of luminosity is expected to come out in the form of X-rays and γ -rays. The spectral shape of this high-frequency emission depends on the particle energy spectrum and on p_n . The nebular radio emission may be detectable, and would provide diagnostic information about the energy spectrum of the e^+e^- pairs in the pulsar wind (E. S. Phinney and C. R. Evans, preprint).

The $H\alpha$ flux integrated over the entire nebula is $3.3 \times 10^{-14} \text{ erg cm}^{-2} \text{ s}^{-1}$. Correcting for extinction ($A_v = 2$; see below) we find an $H\alpha$ output power $L_{H\alpha} \approx 1.7 \times 10^{31} \text{ erg s}^{-1}$, an insignificant fraction of $\dot{E}_R$. Our deep S II narrow-band images and the nebular spectrum (Fig. 2) show that the usual forbidden lines are much weaker than $H\alpha$. Even $H\beta$ is barely detected, if at all. Assuming the little (1.7- σ) blip we see at $\lambda = 4,861 \text{ \AA}$ is a detection of the $H\beta$ line, we find a large Balmer decrement, $I(H\alpha)/I(H\beta) \geq 12$ instead of the usual ratio of about 3. On the basis of total energetics of the pulsar and multicolour data of the companion star, Djorgovski and Evans¹⁷ argue convincingly that the extinction cannot be greater than $A_v = 2$, in which case the intrinsic Balmer decrement is ≥ 6 . If we insist on preserving an intrinsic Balmer decrement of 3, then $A_v \geq 3.7$.

Spectra that are dominated by Balmer line emission are seen in a number of supernova remnants (SNR), such as Tycho¹ (shock speed $v_s \approx 5,600 \text{ km s}^{-1}$) and Cygnus Loop^{27,28} ($v_s \approx 200 \text{ km s}^{-1}$). The Balmer emission is now understood to arise from pre-shock H atoms drifting into the post-shock region² and being excited by collisions with the post-shock hot plasma and by charge exchange with the hot ions. The former gives rise to lines with widths characteristic of the temperature of the pre-shock gas, whereas the latter gives rise to wide profiles representative of the hot post-shock gas. Future high spectral resolution observations may be able to observe these two components. This would be highly desirable as the ratio of the flux

Fig. 2 A calibrated spectrum of the nebula taken with a spectrograph at the Cassegrain focus of the Palomar 60" telescope. Data were obtained on the nights of 17 July and 7 August 1988 and the total integration time is 3 h. The spectrograph slit had a width of 3.8 arcsec aligned along the east-west orientation, and was positioned through the brightest nebula along the southern edge of the nebula. Wavelengths corresponding to the usual emission lines from gaseous nebula are shown. Features marked as NS refer to imperfect subtraction of the night sky lines. The resolution of the spectrograph is $\sim 14 \text{ \AA}$ and the $H\alpha$ is unresolved. $H\beta$ is barely detected, if at all. The spectrum is unreliable for $\lambda < 4,600 \text{ \AA}$ because of the degraded spectrograph focus.



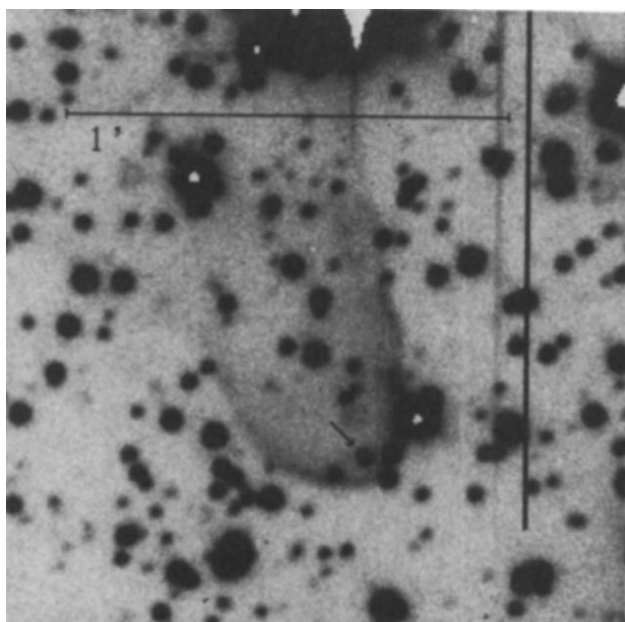


Fig. 3 The stack of exposures of narrow-band (15-Å) $H\alpha$ images taken with the 4-Shooter camera¹⁴ at the prime focus of the Hale 5-m telescope. North is up and east to the left. Each pixel is 0.3 arcsec and the data were obtained under conditions of superb visibility (0.9 arcsec). The star pointed to by the arrow is star X of Kulkarni *et al.*¹⁵, which is now accepted to be the optical counterpart of the pulsar system¹⁶⁻¹⁸. Scale bar indicates 1 arcmin. The dark vertical line to the west of the nebula is a bad column on the CCD in one of the pointings used to construct the image.

in the wide component to that in the narrow component, and the width of the broad component, are excellent measures of v_s . Eventually, the neutral atoms are ionized, at which point they are instantly heated to the post-shock temperatures. In this model, lines from species other than hydrogen are exceedingly weak, primarily because of their low abundance relative to hydrogen.

Given the geometry of PWN1957+20, the shock speed is comparable with the pulsar space motion, so $1 \leq v_7 \leq 2$ where $v_7 = v_s / (100 \text{ km s}^{-1}) (= v_s / 10^7 \text{ cm s}^{-1})$. For shocks in this velocity range, assuming the standard case B (that is, optically thick in all Lyman recombination lines), an $H\alpha$ photon is produced for about every five neutral atoms ionized^{27,29}. The observed $L_{H\alpha}$ implies a particle flux of $F \approx 2.6 \times 10^{43} X^{-1} \text{ atom s}^{-1}$ (where X is the neutral fraction) entering the nebula. Assuming that the nebula has an average angular radius of 15 arcsec (see Fig. 1) or a linear radius of $R \approx 0.08 \text{ pc}$, the particle flux is estimated to be $7.2 \times 10^{42} v_7 n_0 \text{ atom s}^{-1}$. Thus we infer $n_0 v_7 X \approx 3.6$. The warm ionized medium (WIM), the hot ionized medium (HIM) and the warm neutral medium (WNM) have comparable filling factors and collectively pervade the interstellar space³⁰. The pre-shock gas must be the WNM because $X = 0$ for the other two phases. Thus, with $X = 1$, as is appropriate for the WNM³⁰, and $n_0 = 1$ (from our earlier argument), we find $v_7 \approx 3$. With these parameters the power expended in shocking the incident flux of particles is $\frac{1}{2} m_H v_s^2 F \approx 2 \times 10^{33} v_7 \text{ erg s}^{-1}$, which is consistent with the $p dV$ work of $4\pi R^2 v_s p_n \approx 1.1 \times 10^{33} v_7 \text{ erg s}^{-1}$. Thus between 1% and 10% of $\dot{E}_R$ is absorbed in the region outlined by the $H\alpha$ nebula, with the rest of the pulsar wind flowing away from the pulsar.

Models for the Balmer emission give a Balmer decrement of 3 for both case A (that is, optically thin in all Lyman lines) and case B¹, which is smaller than the decrement of ≥ 6 inferred here. The intermediate case of optically thick Lyman- α ($\text{Ly}\alpha$) and $\text{Ly}\beta$ but optically thin $\text{Ly}\gamma$ results in an enhancement of the $n = 3$ level relative to $n = 4$ and hence leads to a decrement

larger than 3 (ref. 2). Indeed, a Balmer decrement of ~ 6 has been observed in fast shocks in the young SNRs, Tycho and SN1006—shocks which are more than one order of magnitude faster than the shock speeds of interest here. In contrast, for the Balmer-dominated emission in Cygnus Loop, where the situation ($v_7 \approx 2$, $n_0 = 2 \text{ cm}^{-3}$, $X \approx 1$) is comparable with PWN1957+20, the Balmer decrement is 3. Clearly, the detailed physics of the shocks in PWN1957+20 is not well understood.

The inferred low ionization of the pre-shocked gas constrains the soft X-ray flux from the synchrotron nebula. If $\frac{1}{4} \dot{E}_R$ comes out as X-rays²² with a power-law spectrum ($I(\nu) \propto \nu^\alpha$; $\alpha < -1$) then the timescale for ionization of a neutral atom at the edge of the nebula, $t_i \approx 1.6 \times 10^6 (E_c / 13.6 \text{ eV})^4 \text{ s}$, where E_c is the low-energy cutoff of the turnover energy. For the ambient medium to be largely unaffected by the ionizing radiation from the PWN, $t_i \gg R / v_p$, which is possible only if $E_c \geq 100 \text{ eV}$, for most reasonable values of α . Thus we predict that the nebular X-ray spectrum must turn over for $E_c \geq 100 \text{ eV}$. Unfortunately, the expected ISM neutral column density ($3 \times 10^{21} \text{ cm}^{-2}$) would make it hard to verify this prediction.

The nebula around PSR1957+20 has very little to do with the ablation of the companion. The ISM flux far exceeds the ablation flux, and besides, the ablated material, being repeatedly shocked, is hot and hence ionized. The physics of the nebula around PSR1957+20 is the same as that used to model the nebula around the young pulsar 1951+32 in CTB80 (ref. 3), the biggest difference being the environment: PSR1957+20 is moving through the standard ISM whereas PSR1951+32 is surrounded by high-density, high-pressure ISM which can confine the pulsar wind more effectively and has led to a rather spectacular PWN. Our failure to detect PWN around other energetic pulsars, including the 1.6-ms pulsar 1937+214, is probably because the ISM in those cases is either of low density or, more probably, is fully ionized.

PWN1957+20 is a new member of a growing list of nebulae around energetic pulsars. A thorough search for other such nebulae will be highly useful because of the diagnostic information that may be gained about the pulsar wind.

We are grateful for stimulating discussions with E. S. Phinney and C. R. Evans, which led to substantial improvements of the paper. S.R.K. is supported partly by a US Presidential Young Investigator's award and the Sloan Foundation. J.J.H. is supported by Caltech's Infrared Processing and Analysis Center.

Received 30 August; accepted 22 September 1988.

- Chevalier, R. & Raymond, J. C. *Astrophys. J.* **225**, L27-L30 (1978).
- Chevalier, R. A., Kirshner, R. P. & Raymond, J. C. *Astrophys. J.* **235**, 186-195 (1980).
- Kulkarni, S. R. *et al. Nature* **331**, 50-53 (1988).
- Blandford, R. D., Ostriker, J. P., Pacini, F. & Rees, M. J. *Astr. Astrophys.* **23**, 145-146 (1973).
- Cohen, N. L., Cotton, W. D., Geldzahler, B. J. & Marcaide, J. M. *Astrophys. J.* **264**, 273-277 (1983).
- Helfand, D. J. in *Supernovae Remnants and their X-ray Emission* (eds Danziger, J. & Gorenstein, P.) 471-484 (Reidel, Dordrecht, 1983).
- Brinkmann, W., Ogelman, H. & Aschenbach, B. *Sp. Sci. Rev.* **40**, 527-529 (1985).
- Alpar, A., Brinkmann, W., Kiziloglu, U., Ogelman, H. & Pines, D. *Astr. Astrophys.* **171**, 101-104 (1987).
- Brinkmann, W. & Ogelman, H. *Astr. Astrophys.* **181**, 71-74 (1987).
- Hester, J. J. & Kulkarni, S. R. *Astrophys. J.* **331**, L121-L125 (1988).
- Shaw, R. A. & Kaler, J. B. *Astrophys. J.* **261**, 510-512 (1982).
- Barker, T. *Astrophys. J.* **219**, 914-930 (1978).
- Fruchter, A. S., Stinebring, D. R. & Taylor, J. H. *Nature* **333**, 237-239 (1988).
- Gunn, J. E. *et al. Opt. Engng* **26**, 779-787 (1987).
- Kulkarni, S. R., Djorgovski, S. & Fruchter, A. S. *Nature* **334**, 504-506 (1988).
- Fruchter, A. S., Gunn, J. E., Lauer, T. R. & Dressler, A. *Nature* **334**, 686-689 (1988).
- Djorgovski, S. & Evans, C. R. *Astrophys. J.* (submitted).
- Van Paradijs, J. *et al. Nature* **334**, 684-686 (1988).
- Cordes, J. M. in *The Birth and Evolution of Neutron Stars* (eds Helfand, D. J. & Huang, J. H.) 34-56 (Reidel, Dordrecht, 1987).
- Cheng, A. F. & Helfand, D. J. *Astrophys. J.* **271**, 271-282 (1983).
- Cheng, A. F. *Astrophys. J.* **275**, 790-801 (1983).
- Arons, J. *Nature* **302**, 301-305 (1983).
- Phinney, E. S., Evans, C. R., Blandford, R. D. & Kulkarni, S. R. *Nature* **333**, 832-834 (1988).
- Kluźniak, W., Ruderman, M., Shaham, J. & Tavani, M. *Nature* **334**, 225-227 (1988).
- Van den Heuvel, E. P. J. & van Paradijs, J. *Nature* **334**, 227-235 (1988).
- Kulkarni, S. R. in *The Physics of Black Holes and Neutron Stars* (ed. Tanaka, Y.) (in the press).
- Raymond, J. C., Davis, M., Gull, T. R. & Parker, R. A. *Astrophys. J.* **238**, L21-L25 (1980).
- Raymond, J. C., Blair, W. P., Fesen, R. A. & Gull, T. R. *Astrophys. J.* **275**, 636-644 (1983).
- Cox, D. P. & Raymond, J. C. *Astrophys. J.* **298**, 651-659 (1985).
- Kulkarni, S. R. & Heiles, C. in *Galactic and Extragalactic Astronomy* (eds Verschuur, G. L. & Kellerman, K. I.) 95-151 (Springer, New York, 1988).

Circumstellar matter of SN1987A and soft X-ray emission

K. Masai*, S. Hayakawa†, H. Inoue‡, H. Itoh§ & K. Nomoto||

* Institute of Plasma Physics and † Department of Astrophysics, Nagoya University, Nagoya 464-01, Japan

‡ Institute of Space and Astronautical Science, Sagami-hara 229, Japan

§ Department of Astronomy, Kyoto University, Kyoto 606, Japan

|| Department of Earth Science and Astronomy, University of Tokyo, Tokyo 153, Japan

X-ray emission from supernova 1987A has been observed continually by the Ginga satellite since July 1987. The observed X-ray spectrum^{1,2} has two components: hard X-rays, above 15 keV in energy, have a nearly flat spectrum, and are thought to be Compton-degraded γ -rays from the decay of ^{56}Co (ref. 3); the soft component (<15 keV) shows a thermal bremsstrahlung spectrum with a temperature of 10 to 12 keV, and may be due to thermal emission from shock-heated ejecta⁴. In January 1988, X-ray emission in the range 6–16 keV flared up, on a timescale of a few weeks, to about three times the level of the previous three months⁵. Here we interpret the flaring as the result of interaction of the supernova ejecta with pre-existing circumstellar matter, whose density distribution was peaked because material blown off from the progenitor star ran into slower-moving material ejected during an earlier stage of the star's life.

Between early September and late December 1987, the intensity of the hard component of the X-ray emission from SN1987A was constant within statistical errors, whereas that of the soft X-rays fluctuated by a factor of about two on a timescale of a month. In January 1988 the intensity increased steeply over two weeks and then declined over a month to the preceding level. The intensities of the soft and hard components at the peak of this flare were about three and two times the average pre-flare intensities, respectively. If the intensity and spectrum of the Compton-degraded X-rays are essentially constant, as a result of the mixing of ^{56}Co into the outer region^{6,7}, the flare can be explained as a strong enhancement of thermal emission associated with an increase in the emission measure (by a factor of three) and in the electron temperature.

In an earlier paper⁸ we predicted that the collision of expanding ejecta with circumstellar matter (CSM) could produce thermal X-rays. By comparing this model with the X-ray data obtained in September 1987, we demonstrated that the CSM could be formed by the stellar wind in the final phase of a red supergiant or in the transition phase to the blue supergiant that underwent the supernova explosion⁴. Continuous study of the supernova by Ginga has provided further information on the density structure of the CSM where it interacts with the ejecta. Following previous calculations^{4,8}, we construct here a model of CSM that can account for the observed soft X-rays and we suggest that the interaction region may be considerably obscured by the ejecta.

As indicated from recent ultraviolet observations⁹, the progenitor of SN1987A had evolved into a red supergiant and then contracted to a blue supergiant¹⁰. In the red phase the stellar wind is ejected at a speed of $\sim 10 \text{ km s}^{-1}$. As the star evolves to the blue phase, the wind speed increases and the mass loss rate decreases. Matter ejected later collides with that ejected earlier. This is similar to the situation in a planetary nebula, where gas slowly ejected by a red giant collides with a fast wind from a shrunken hot star¹¹. The colliding materials are compressed by shocks to form a high-density shell. This massive shell is subject to a Rayleigh-Taylor instability and breaks up into fragments, some of which remain behind the slowly expanding shell. A dense circumstellar shell may also be formed by a sporadic mass

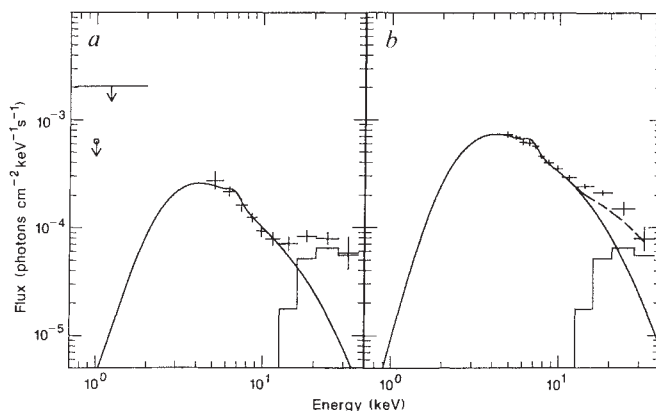


Fig. 1 Calculated spectra at *a*, ~ 200 d and *b*, ~ 330 d, which are convoluted to an energy resolution of $0.5 (E/\text{keV})^{1/2} \text{ keV}$ (full width at half maximum) at a photon energy E . The crosses represent the observed spectra²: *a*, pre-flare average in 1987; *b*, the flare in January 1988. The upper limits at 0.75–2 keV and at 1 keV, determined from rocket observations^{14,15}, are also shown as symbols with arrows in *a*. Histograms represent the residuals of the observed flux above the thermal spectrum in *a*. The dashed line in *b* represents the sum of thermal emission and the residuals.

ejection shortly before the explosion, as proposed for SN1984E¹².

The CSM, whether formed by the interaction of fast and slow winds or by non-steady mass ejection, will be inhomogeneous and lacking in spherical symmetry. Moreover, the distribution of supernova ejecta is likely to be asymmetric, as suggested from recent speckle observations¹³. Inhomogeneity in the CSM may be responsible for observed fluctuations of the soft X-ray intensity, and if the soft X-ray emission is predominantly on the far side, absorption by intervening matter explains the low X-ray intensity below 2 keV.

We can account for the observed X-ray features, using a simple model of the CSM in which we assume that the X-ray-emitting surface is located only on the far side and covers <30% of the total spherical area. Such a model incorporates both the reduction of the ejecta surface, which is responsible for X-ray emission, and the absorption effect, which requires a hydrogen column density $> 1 \times 10^{22} \text{ cm}^{-2}$, and is consistent with both the spectra observed by Ginga⁵ and the upper limits on emission imposed from rocket experiments^{14,15}. Hence, in this model the X-rays emitted from the central part of the interaction region are totally obscured, and those from the peripheral part are absorbed by the ejecta. A small emission area and the obscuration of X-rays are compensated by a high emission measure. The initial density distribution of the CSM is assumed to be $2.4 \times 10^5 (r/(1.0 \times 10^{16} \text{ cm}))^{-2} \text{ AMU cm}^{-3}$ for $1.0 \times 10^{16} \text{ cm} \leq r \leq 3.2 \times 10^{16} \text{ cm}$, where r is the distance from the supernova. A dense shell is formed at $r \approx 3.2 \times 10^{16} \text{ cm}$ with a density of $3 \times 10^5 \text{ AMU cm}^{-3}$ and a thickness of $9.4 \times 10^{14} \text{ cm}$, which covers $\approx 16\%$ of the spherical area on the far side. The presence of such a thin shell is crucial for explaining the X-ray flare. This model is much better able to reproduce the observed spectra⁵ than is a model with a symmetrical shell, incorporating interstellar absorption that corresponds to a hydrogen column density as large as 10^{23} cm^{-2} .

The ejecta model used here is based on a hydrodynamic calculation for modelling the optical light curve¹⁶. Metal content of the envelope is one-third of the solar value. Relative abundances in the CSM are assumed to be the same as those in the envelope, and 13 elements (H, He, C, N, O, Ne, Mg, Si, S, Ar, Ca, Fe and Ni) are taken into account in calculations of emission and absorption. We approximate the initial density distribution in the outer envelope by a power law, $r^{-9.75}$. We calculate the energy flux and spectrum at the appropriate distance of 52 kpc

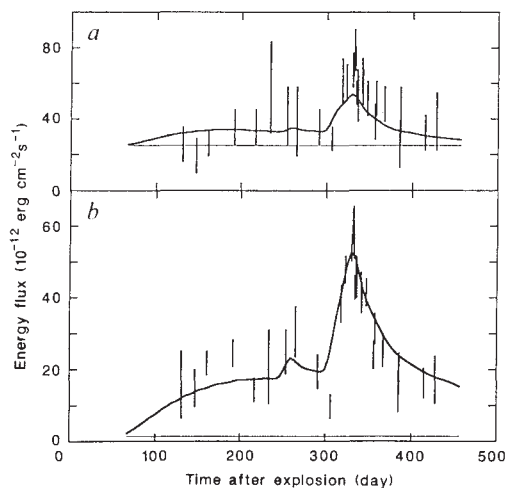


Fig. 2 Light curves (heavy lines) in the energy ranges 16–28 keV (a) and 6–16 keV (b). The horizontal thin lines represent the constant fluxes of Compton-degraded X-rays and provide the base lines for the light curves. The bars represent the observed data, converted from counts s^{-1} on the LAC (Large Area Counter)⁵ to $erg\ cm^{-2}\ s^{-1}$ using ref. 25.

from SN1987A, with interstellar absorption equivalent to a neutral hydrogen column density of $2 \times 10^{21}\ cm^{-2}$ and solar abundances of elements.

With this model, the X-ray brightening starts after ~ 70 days, as observed. Thermal X-ray emission is dominated by that from the reverse-shocked ejecta unless the 'swept-up' mass of the CSM becomes larger than the ejecta mass. Thus the light curve is determined primarily by the time evolution of the outer envelope of ejecta. The soft X-rays observed since July 1987 are due mainly to the reverse shock arising from the collision with the inner CSM. The outwardly propagating blast wave hits the outer dense shell and enhances the soft X-rays. This blast-shock precursor is followed within a few months by a larger flare from propagation of the reverse shock into the ejecta. The flare observed in January 1988 thus results from re-heating of the envelope by the second reverse shock, which rises, in turn, from the collision with the dense shell.

This model gives an X-ray spectrum at ~ 200 d that is dominated by free-free (thermal bremsstrahlung) emission of temperature ~ 12 keV and is consistent with observations^{5,14,15} (Fig. 1a). The contribution from the Compton-degraded X-rays is represented by the residual of the observed flux⁵ above the thermal spectrum, as shown by histograms in Fig. 1. The spectrum calculated at ~ 330 d is presented in Fig. 1b, along with the spectrum observed for the flare in January 1988. The electrons are re-heated to ~ 17 keV by the reverse shock. For the flare spectrum we predict a blend of iron $K\alpha$ -lines (that is, a complex of unresolved lines) at a photon energy ~ 6.8 keV, with an equivalent width of ~ 250 eV, which are consistent with those observed⁵. In Fig. 1b, the sum of thermal emission and the residuals in Fig. 1a is shown as a dashed line. The excess observed over the thermal spectrum at 15–25 keV is too large to be explained in terms of Compton-degraded X-rays if their intensity is constant after 200 d, so that an increase in their intensity may be implied and/or a possible contribution from a suprathermal component. If electrons as well as ions are heated at the collisionless shock front, the average electron temperature could be increased to ~ 37 keV, but this would not be sufficient to account for the excess around ~ 20 keV.

Radio flux from free-free emission is lower than $22\ \mu Jy$ at 1.4 GHz, which is well below the upper limit of observation. Synchrotron radio emission associated with shocks¹⁷ is subject to strong free-free absorption because the optical depth is much larger than unity at 43 GHz.

The light curves calculated in two energy ranges are shown in Fig. 2, in which the residual fluxes in Fig. 1a ($\sim 1 \times 10^{-12}\ erg\ cm^{-2}\ s^{-1}$ at 6–16 keV and $\sim 25 \times 10^{-12}\ erg\ cm^{-2}\ s^{-1}$ at 16–28 keV) are added to the flux of thermal emission. The light curves show a small enhancement at ~ 260 d, which arises from the collision of the expanding blast wave with the dense shell. At ~ 330 d, the inwardly-propagating reverse shock hits the ejecta, causing a flare. X-rays from the shocked ejecta decay rapidly by adiabatic expansion. The light curve observed for 6–16 keV is fairly well reproduced by our model. The 16–28 keV light curve can be explained in terms of thermal and Compton-degraded X-rays, except around the peak of the flare. The somewhat larger deviation from our predictions for this latter light curve could imply a steady increase in Compton-degraded hard X-rays, as indicated by a recent observation of 30–110-keV X-rays¹⁸.

As shown in Figs 1 and 2, thermal X-rays produced by the interaction of supernova ejecta with the CSM explain the gross features of soft X-ray emission from SN1987A, but production of 15–25-keV X-rays requires further investigation. The CSM model discussed here has implications for the evolution of the supernova progenitor. The matter in the inner CSM and the dense shell is considered to be ejected after the onset of contraction towards the blue stage. The time elapsed since the red supergiant wind ceased is estimated, from analysis of near-infrared speckle data¹⁹ and optical O III line observations²⁰ to be $\sim 6,000$ yr. It also takes $\sim 6,000$ yr for a star with a $6\ M_{\odot}$ helium core to contract from the red to the blue²¹. If the CSM were formed originally by a spherically symmetric stellar wind and extended to the inner edge of the red supergiant dust shell at $\sim 3 \times 10^{17}$ cm (ref. 19) with the density distribution assumed here, its mass and the average mass loss rate would have been $\sim 7 \times 10^{-2}\ M_{\odot}$ and $\sim 10^{-5}\ M_{\odot}\ yr^{-1}$, respectively. Relative to previous studies^{17,22} this rate seems to be too large. If the CSM terminates at a few times 10^{16} cm, its origin could be attributed to a sporadic mass ejection ≤ 100 yr before the explosion. But in view of the uncertainty in the mass loss history of the star evolving from a red to a blue supergiant it is necessary to investigate the time variations of mass loss rate and wind speed, interaction between winds, and wind geometry, to understand the CSM structure deduced here. The existence of a nearby dense cloud is also proposed by Hillebrandt *et al.*²³, who, in an attempt to explain the 'mystery spot', predicted an X-ray enhancement about one year after the explosion.

For the CSM outside the dense shell, we assume here that a stellar wind blown at a rate of $10^{-6}\ M_{\odot}\ yr^{-1}$ for 6,000 years expands to $\sim 3 \times 10^{17}$ cm. Then the intensity of thermal X-rays will decline slowly until the supernova ejecta expands to interact with the dense matter ejected in the red supergiant phase. Some other dense clouds formed in the red-to-blue transition may be embedded in the CSM. If this is the case, X-ray enhancements similar to the flare in January 1988 will be observed.

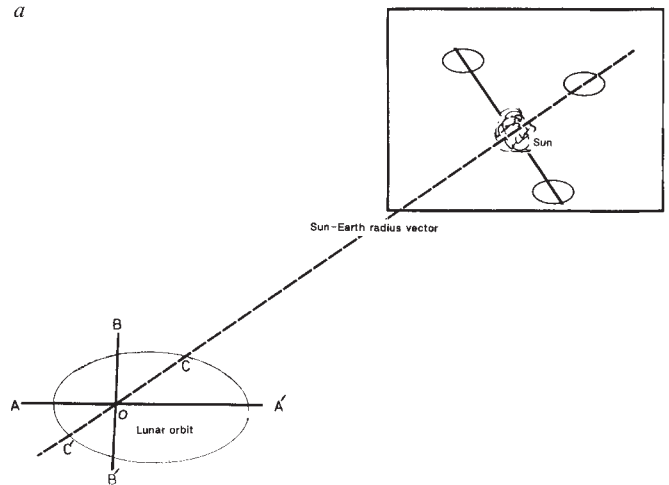
Finally we comment on an alternative interpretation of the soft X-rays. Bandiera *et al.*²⁴ proposed that soft X-rays could be produced by a central pulsar. Because of absorption by the surrounding ejecta, the spectrum should exhibit reduced absorption as the ejecta expands. No appreciable softening of the spectrum with time is observed, which disfavors the pulsar hypothesis. Moreover, the presence of an iron emission line restricts the possible contribution of pulsar X-rays.

Received 13 June; accepted 16 September 1988.

- Dotani, T. *et al. Nature* **330**, 230–232 (1987).
- Sunyaev, R. *et al. Nature* **330**, 227–229 (1987).
- Itoh, M., Kumagai, S., Shigeyama, T., Nomoto, K. & Nishimura, J. *Nature* **330**, 233–235 (1987).
- Masai, K., Hayakawa, S., Itoh, H. & Nomoto, K. *Nature* **330**, 235–236 (1987); corrigendum **330**, 588 (1987).
- Tanaka, Y. in *Atmospheric Diagnostics of Stellar Evolution* IAU Colloquium 108 (ed. Nomoto, K.) 399–405 (Springer, Berlin, 1988).
- Ebisuzaki, T. & Shibazaki, N. *Astrophys. J.* **327**, L5–L8 (1988).
- Kumagai, S., Itoh, M., Shigeyama, T., Nomoto, K. & Nishimura, J. *Astr. Astrophys.* **197**, L7–L10 (1988).

8. Itoh, H., Hayakawa, S., Masai, K. & Nomoto, K. *Publ. astr. Soc. Japan* **39**, 529–537 (1987).
9. Panagia, N. *et al.* *IAU Circ.* no. 4514 (1987).
10. Saio, H., Kato, M. & Nomoto, K. *Astrophys. J.* **331**, (in the press).
11. Kwok, S. in *Mass Outflows from Stars and Galactic Nuclei* (eds Bianchi, L. & Gilmozzi, R.) 123–136 (Kluwer Academic Publishers, 1988).
12. Gaskell, C. M. & Keel, W. C. in *Supernova 1987A in the Large Magellanic Cloud* 4th George Mason Univ. Workshop (eds Kafatos, M. & Michalitsianos, A. G.) 13–15 (Cambridge Univ. Press, 1988).
13. Karovska, M. *et al.* *IAU Circ.* no. 4604 (1988).
14. Aschenbach, B. *et al.* *Nature* **330**, 232–233 (1987).
15. Burrows, D., Nousek, J. & Garmire, G. *IAU Circ.* no. 4494 (1987).
16. Shigeyama, T., Nomoto, K. & Hashimoto, M. *Astr. Astrophys.* **196**, 141–151 (1988).
17. Chevalier, R. A. *Nature* **332**, 514–516 (1988).
18. Ubertini, R. *et al.* *IAU Circ.* no. 4590 (1988).
19. Chalabae, A. A., Perrier, C. & Mariotti, J. M. in *Supernova 1987A in the Large Magellanic Cloud* 4th George Mason Univ. Workshop (eds Kafatos, M. & Michalitsianos, A. G.) 236–239 (Cambridge Univ. Press, 1988).
20. Wampler, E. J. & Richichi, A. *The Messenger (ESO)* **52**, 14–16 (1988).
21. Saio, H., Nomoto, K. & Kato, M. *Nature* **334**, 508–510 (1988).
22. de Jager, C., Nieuwenhuijzen, H. & van der Hucht, K. A. *Astr. Astrophys. Suppl.* **72**, 259–289 (1988).
23. Hillebrandt, W., Höflich, P., Schmidt, H. V. & Truran, J. W. *Astr. Astrophys.* **186**, L9–L10 (1987).
24. Bandiera, R., Pacini, F. & Salvati, M. *Nature* **332**, 418–419 (1987).
25. Makino, F. *IAU Circ.* no. 4530/4532 (1988).

a



b

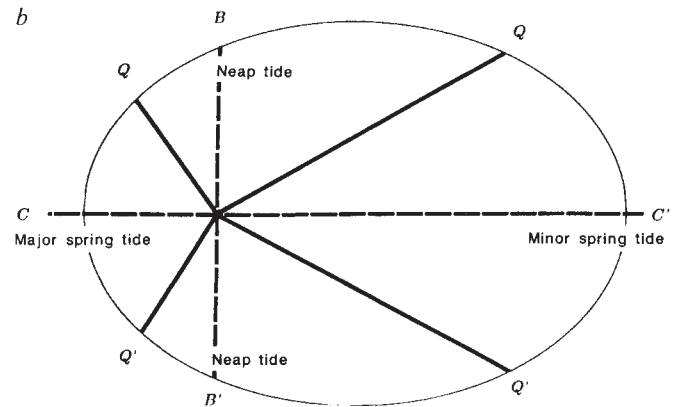


Fig. 1 a, The lunar orbit (detail) and the motion of the orbit about the Sun. The Earth is at the focus O. Partial lunar orbital tidal amplitude maxima occur when the Moon is aligned along CC', the radius vector from the Sun (new or full moon), but the two monthly maxima are generally unequal except when the radius vector to the Sun is aligned with the latus rectum, BB'. Partial lunar orbital maxima peak for alignment of the radius vector with AA'. Further effects from non-coplanarity of orbits are not shown. b, Configuration of the lunar orbit diagramming orbital arcs when sedimentation is 'on'. These are QCQ' and QC'Q'—periselene and aposelene. The alternate arcs, Q'B'Q' and QBQ, correspond to times given in Fig. 2 when tidal cutoff occurs. The configuration shown here corresponds to the semi-annual maxima when the radius vector from the Sun is aligned with the major axis, CC'. The alternative is when aligned along the latus rectum (see Fig. 2).

The lunar orbit in the late Precambrian and the Elatina sandstone laminae

C. P. Sonett*, S. A. Finney & C. R. Williams

Department of Planetary Sciences and *Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA

The discovery of layered sediments of the Elatina Formation of South Australia has attracted intense speculation regarding their origin^{1,2}. The formation 10-m layer of graded sandstone, consists of a pattern of light bands of periodically varying thickness³. A sequence of dark bands separates, on average, 11.6 light bands; the spacing of these dark bands exhibits a rich and complex spectrum. Previously, the periodicity has been attributed to the sunspot cycle and unassigned but putative solar periods. Prompted by a letter from G. E. Williams, summarizing new field data from the Adelaide region, South Australia, which imply that the Elatina varves may be tidal, we re-examined the dark-band spectrum and propose a luni-solar tidal interaction model as the source of the laminae.

The lunar tidal disturbance potential ϕ_m induces a gravitational quadrupole moment in the Earth which has prolate spheroidal symmetry about the Earth–Moon radius vector⁴. The eccentricity of the lunar orbit introduces a monthly amplitude modulation into the tidal signal (Fig. 1). Because of the additional solar tidal potential, ϕ_s , the monthly modulation is divided into two partial lunar orbit (PLO) envelope maxima of generally differing amplitudes but which become equal twice a year ($OC = OC'$, Fig. 1) when the solar radius vector lies along the latus rectum, BOB'.

At some point P on the Earth's surface, the disturbance potential (tide raised on Earth by the Moon or Sun) can be written as

$$\phi_{m,s} = -GM_{m,s}a^2/2r_{m,s}^3[3\cos^2\psi_{m,s} - 1] \quad (1)$$

where

$$\cos\psi = \sin\theta\cos\lambda \quad (1a)$$

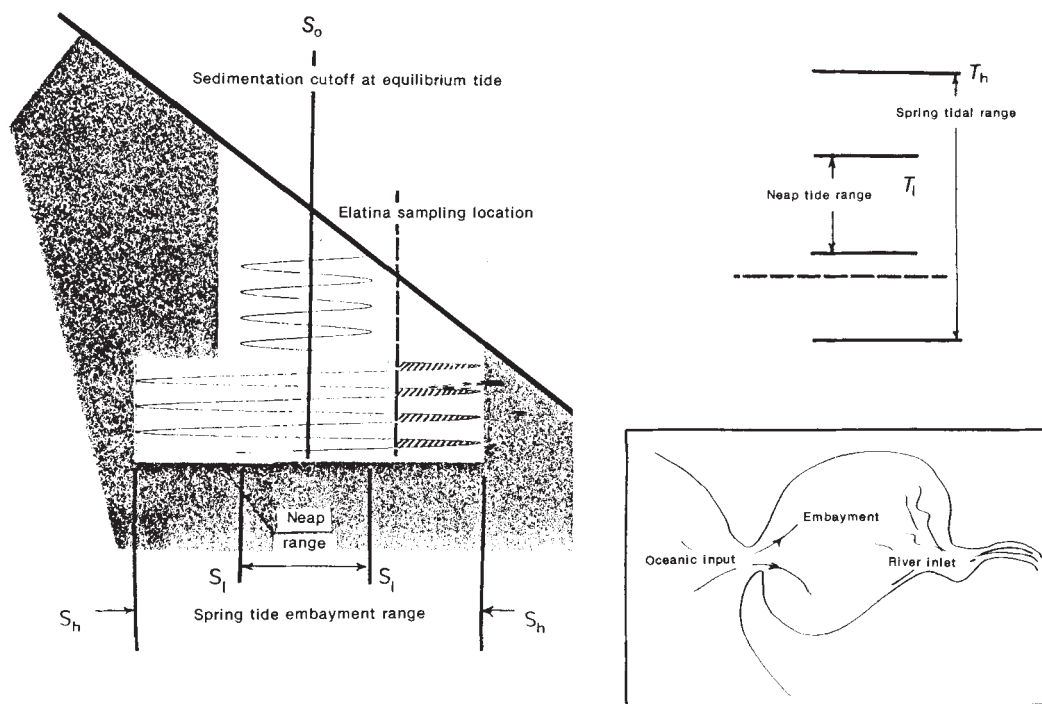
θ being the colatitude of P on the Earth's surface, $\lambda = \omega_e t$ is the time-dependent longitude measured eastward from the Earth–Moon or Earth–Sun radius vector where ω_e is the rotational angular velocity of the Earth and t is time, G is the gravitational constant, subscripts m and s represent the lunar and solar case respectively, M is mass, a is the mean radius of the Earth, and r is the separation between the Earth and the Sun or Moon. The total tidal disturbance potential (Moon and Sun) is

$$\phi_T = \phi_m + \phi_s \quad (2)$$

The equilibrium tide (inertial forces neglected), corresponding to ϕ_m is raised semi-diurnally on a rotating Earth where the spin axis is normal to the plane of the Moon's orbit. Now, with the lunar orbit inclination to the Earth's Equator of $23^\circ \pm 9'$, a strong latitude-dependent inequality in the two semi-diurnal tides⁵ (the so-called 'tidal inequality').

In the tidal model, the varve-generating lacustrine scenario of Williams^{1,6} is replaced by a river-fed lagoon or embayment (Fig. 2). The scale of the basin is taken to be small compared to water-wave speeds and the basin adjusts to changes in water level in a time short compared to the diurnal tidal period. A long fetch is ruled out, on the basis of sedimentary evidence, as transmission of coastal oceanic wave motions; and storm surges must have been unimportant, lest over the ~ 53 yr of deposition (based on our reckoning of 26 date bands per year,

Fig. 2 Schematic of the (linear) tidal modulator. The exaggerated slope of the bottom of the embayment performs a linear operation on the tidal range. Because the hypothesized Elatina site is offset from the mean equilibrium tidal position, there is variable sampling of the daily range of the embayment range defined by the point of significant energy exchange between the lagoon water and the river flow (not shown) entering from the left.



as discussed later) the delicate layering would have been compromised.

As the river enters the embayment carrying sediments it is slowed by interactive contact with basin waters, at which time the flow begins, losing some of its carrying power and larger particles settle out closer to the river/lagoon interface, whereas smaller particles are carried further into the basin by the vestiges of the river current. At high tide only smaller fines, clays for example, are deposited at the Elatina site. At low tide, larger particles are laid down at the site. Thus, varve-like deposits are produced on a timescale of two light bands per day. In a simplified system such as this, it is expected that fine sediments correspond to the highest tide, then coarsen upwards from the high tide position to the low tide position, followed by fining upwards to the next point of high tide⁷.

In Fig. 2, S represents the farthest point where the finest riverborne sediments are deposited; to the right of S , sedimentation is negligible. At equilibrium, S is located at S_0 . The swing in tidal amplitude causes the position of S to vary as shown. The Elatina site will accumulate sediment as long as the tide remains below a certain level. On a daily basis, two low tides will result in two deposition periods (Fig. 2). For small tidal variability between high and low tide, coarse clastic deposition can be interrupted for days although the very fine grains suspended in the lagoon water settle to the bottom, forming a dark band. These provide an unbroken record of PLO tides as one dark band is deposited at each low in the tidal period. It is the expected offset of the Elatina location which results in a sedimentation cutoff and reduction of light-band count below the expected PLO value. Thus we cannot recover an unbroken record of the daily laminations (light bands); they are an incomplete set. Frequency modulation (FM) reported previously⁸ is explained by variations in the tidal cutoff changing the number of semi-diurnal laminae captured during one PLO. Thus a natural FM takes place, corrupting the number of light laminae expected in a PLO, producing an unreliable clock not used here.

The distance between adjacent dark bands is a complicated function of the tidal range and the length of time for which the water-level was low. The existence of a tidal cutoff is inferred

directly from the ~ 11.6 light laminae per dark band⁹. If these represent semi-diurnal layers, then a full complement of 31 layers should be deposited in a PLO, whereas less than half this is seen. Even a full tidal inequality would lead to deposition of 15 laminae. It is possible that sufficient light bands actually are present to define a daily tide.

The relation between dark-band thickness and the number of light laminae per dark band which mimics the sunspot index-length of solar activity cycle is explained as follows. Figure 1b shows the influence upon tidal disturbance potential maxima about periselene and aposelene (orbital arcs QCQ' and $QC'Q'$). Whereas the spring tide conditions prevail for both, the tide at QCQ' is larger than its counterpart at $QC'Q'$ because of the Moon's proximity. The passage of time, however, through arc QCQ' is less than that through $QC'Q'$. Thus, sedimentation from QCQ' will show a greater deposition rate and thus a wider dark band whereas the number of days (light band pairs) will be less than at $QC'Q'$.

On dynamical grounds, the PLO pairs of tides defined within a lunar month should generally be unequal in amplitude (and therefore define unequal dark band widths) except for twice per year when they are equal and exchange order. Furthermore, taking account of the advance of perigee of the lunar orbit means that the tidal year, defined by the apsidal advance, is longer than the solar year. The unequal dark banding and its cyclic interchange twice in a period significantly longer than the solar year are potential confirmation of the luni-solar interaction.

The tidal period τ_t is defined by twice the intervals between reversals in the order of the thin and thick dark bands ('phase flips') and results from the annual luni-solar tidal potential variation. Although absent in the dark-band power spectrum, this period is identified by examination of the difference in width of the thin and thick bands⁶, yielding $\tau_t = 29.2$ PLO with estimated error ≤ 0.06 . With $\tau_s = 26.25$ PLO per solar year, this tidal period corresponds to $\tau_s = 1.112$ solar years. This 11.2% difference between the observed tidal year and the solar year is explained by the rotation of the lunar orbital periapsis. The apsidal rotation thus has a sidereal period of 9.43 ± 0.14 years.

Table 1 Estimated orbital parameters of the late Precambrian Earth–Moon system

Parameter	Numerical value
Synodic lunar frequency*	$3.8095 \pm 0.003 \times 10^{-2}$ yr per PLO†
Sidereal lunar orbits	14.12 ± 0.01 yr ⁻¹
Terrestrial angular velocity	$7.94 \pm 0.01 \times 10^{-5}$ rad s ⁻¹
Lunar semi-major axis	$3.678 \pm 0.002 \times 10^5$ km
Mean Earth–Moon separation	$54.53 \pm 0.03 R_e$
Lunar orbit apsidal period	9.71 ± 0.01 yr
Length of day	21.98 ± 0.03 h
Days per year	398.8 ± 0.5
Mean rate of retreat of the Moon‡	2.02 ± 0.02 cm yr ⁻¹
Terrestrial frictional dissipation rate‡	$1.785 \pm 0.03 \times 10^{12}$ W

Excluding solar tidal contribution to angular momentum and energy flux; all errors are based upon worst-case combinations of individual 2σ errors of the lunar orbital and tidal frequency.

* Bayesian estimate of spectral line position.

† PLO denotes partial lunar orbits.

‡ Based upon Precambrian age of 6.8×10^8 yr.

The dark-band spectrum consists of a sequence of very evenly spaced harmonic spectral features^{6,9}. The fundamental appears to be associated with a fine structure of three closely spaced but well-defined narrow lines with periods of ~ 1 year. Identification of these line frequencies with the solar year yields a range of values for all orbital parameters except eccentricity. Here we have chosen the strongest of the three lines to represent the yearly (seasonal) forcing. Errors quoted are based upon the two-standard-deviation estimate for the accuracy of the spectral line used for Table 1, at a frequency $f = 0.038095$ years per PLO. Those parameters also dependent upon lunar tidal theory are subject to larger, but uncertain errors. The estimate of the apsidal period is based upon Table 1 of ref. 10, from which the first seven terms are used to guarantee convergence. Nevertheless, errors are still present as the theory makes significant approximations¹¹.

Estimates of the apsidal period of the Precambrian Moon vary from 9.92 to 9.26 yr depending upon the choice of spectral line, and the mean retreat rate of the Moon ranges from 2.47 to 1.00 cm per year using the extrema of the three spectral lines grouped about one year, combined with two-standard-deviation error limits. All parameters reported here are based upon the assumption that the major power at $f = 0.038095$ PLO is the proper choice for the solar year measured in PLOs. The choice of $f = 0.038095$ years per PLO is supported by the difference in apsidal period determined from the sediments and that derived using lunar theory, being a minimum when using this value for f . But we regard the values given here for the Precambrian lunar orbit as tentative.

The solar tidal contribution is detected by an overall amplitude modulation of the dark-band system, seen in both the time sequence and the power spectrum. From the power spectrum of the dark bands it is deduced that the amplitude modulation repeats with a period of 26.25 PLO. Thus there are 14.12 sidereal lunar months. The semi-major axis $a = [GM_e(\tau/2\pi)^2]^{1/3}$, where τ is the lunar orbital period, is estimated to have been $3.6780 \pm 0.002 \times 10^{10}$ cm, a decrease of 1.374×10^9 cm relative to the present-day value. It follows that the lunar distance was $54.53 R_e$ and the mean rate of retreat of the Moon up to the present has been 2.02 ± 0.02 cm per yr.

All error estimates are 2σ worst-case combinations from the two periods found from the data and which form the basic input parameters. The late Precambrian lunar orbit is assumed to be inclined at the present value of 5.15° to the plane of the ecliptic, but may have been less or more. The nodal regression does not enter into the arguments given here, but is expected to affect the tidal inequality. The tidal year, τ_t , is confirmed by the alternation of dark bands, and the presence of the shorter solar year τ_s depends to a large extent on the variation of the solar gravitational potential resulting from the Earth's orbital eccentricity, although a climatic contribution from variable riverine

flow may exist. The late Precambrian Earth–Moon distance compares favourably with the present-epoch Apollo laser value of $\dot{r} = 3.47$ cm per yr (ref. 12). The small value we find for the late Precambrian agrees with lessened tidal friction at that time and a recent increase in angular momentum transfer through oceanic resonance¹³.

Most ancient length-of-day estimates cited¹⁴, in particular the modern eclipse¹⁵ and current Apollo laser value, are marginally higher than those found here, implying slightly larger tidal friction, but they are all consistent with increased tidal friction over the past 6.66×10^9 yr although there is a wide spread in determinations of the tidal lunar acceleration (see Table 1 of Calame and Mulholland¹⁶).

Estimates of the concomitant terrestrial tidal dissipation vary but all seem low, being in the range of 10^9 W for the deep ocean¹⁷, 2.2×10^{11} W for the atmosphere, and $3\text{--}5 \times 10^{12}$ W overall^{18–20}. The latter value (presumably modern), supports the idea of increased dissipation with time and a comparison with our values is consistent with the view that the break-up of Pangea was associated with the later development of shallow seas and higher friction in the Palaeozoic, but the larger value from the Apollo laser supports the calculations of Hansen¹³ and the importance of oceanic resonances in the present epoch. It is uncertain whether Australia was equatorial¹⁶ or southerly²¹ in the late Precambrian and this would have an as yet unresolved effect upon any tidal inequality at the Elatina site.

Long-term averages of the energetics of the Earth–Moon system can be obtained from the redistribution in system energy. The current value of orbital energy $E = -GM_e/2a$ is -3.839×10^{35} erg; the corresponding ancient value was -3.983×10^{35} erg and thus $\Delta E = 1.083 \times 10^{33}$ erg. This lessened orbital energy must appear as increased rotational energy of the Earth and increased tidal friction. Partitioning energy, while conserving angular momentum yields values for the ancient Earth's rotational energy, velocity and the frictional heating loss $\dot{H} = 9.42 \times 10^{-9}$ J gm-yr⁻¹ or $1.78 \pm 0.02 \times 10^{12}$ W. The heating loss in the Earth is comparable to 2.1×10^{-7} J gm-yr⁻¹ for ordinary chondrites and a range of terrestrial rocks⁴.

We thank J. G. Williams, R. Greenberg and G. Ojakangas for carefully reading the manuscript, to R. Y. Anderson and J. Melosh for helpful discussions, and to G. Bretthorst for his aid in illuminating his Bayesian estimator. This research was supported by the NSF's Solar-Terrestrial program. We understand that a paper from G. E. Williams, containing his new data and model will be appearing soon.

Received 21 April; accepted 23 August 1988.

- Williams, G. E. *Nature* **291**, 624–628 (1981).
- Williams, G. E. *Aust. J. Phys.* **38**, 1027–1043 (1985).
- Finney, S. A., Williams, C. R. & Sonett, C. P. *Proc. 19th Lunar and Planetary Sci. Conf.* 331–332 (Lunar & Planetary Inst., Houston, 1988).
- Stacey, F. D. *Physics of the Earth* (Wiley, New York, 1977).
- Defant, A. *Physical Oceanography*, Vol. 2 (Pergamon, Oxford, 1961).
- Williams, G. E. *Aust. J. Phys.* **38**, 1027–1043 (1985).
- Prior, D. B., Bornhold, Brian D., Wiseman, W. J. Jr & Lower, D. R. *Science* **237**, 1330–1333 (1987).
- Sonett, C. P. & Williams, G. E. *Sol. Phys.* **110**, 397–410 (1987).
- Williams, G. E. & Sonett, C. P. *Nature* **318**, 523–527 (1985).
- Brouwer, D. & Clemence, G. M. *Methods of Celestial Mechanics* (Academic, New York, 1961).
- Danby, J. M. A. *Fundamentals of Celestial Mechanics* (MacMillan, New York, 1962).
- Newhall, X. X., Williams, J. G. & Dicke, J. O. in *Rotation and Reference Frames for Geodesy and Geodynamics* (eds Babcock, A. K. & Wilkins, G. A.) 159–164 (Reidel, Dordrecht, 1988).
- Hansen, K. S. *Rev. Geophys. Spa. Res.* **20**, 457–480 (1982).
- Creer, K. M. in *Growth Rhythms and The History of The Earth's Rotation* (eds Rosenberg, G. D. & Runcorn, S. K.) 293–318 (Wiley, New York, 1975).
- Muller, M. & Stephenson, F. R. in *Growth Rhythms and The History of The Earth's Rotation* (eds Rosenberg, G. D. & Runcorn, S. K.) 459–534 (Wiley, New York, 1975).
- Calame, O. & Mulholland, J. D. in *Tidal Friction and The Earth's Rotation* (eds Brosche, P. & Sünderman, J.) 43–54 (Springer, Berlin, 1978).
- Jeffreys, H. *The Earth* (Cambridge University Press, 1970).
- Tarling, D. H. in *Growth Rhythms and The History of The Earth's Rotation* (eds Rosenberg, G. D. & Runcorn, S. K.) 397–426 (Wiley, New York, 1975).
- Morrison, L. V. in *Growth Rhythms and The History of The Earth's Rotation* (eds Rosenberg, G. D. & Runcorn, S. K.) 445–458 (Wiley, New York, 1975).
- Miller, G. R. *J. geophys. Res.* **485**–489 (1966).
- Smith, A. G. & Hallam, A. *Nature* **225**, 139–144 (1970).

Palaeoecological evidence for early aragonite dissolution in ancient calcite seas

T. J. Palmer*, J. D. Hudson† & M. A. Wilson‡

* Department of Geology, University College of Wales, Aberystwyth, Dyfed SY23 3DB, Wales, UK

† Department of Geology, University of Leicester, Leicestershire LE1 7RH, UK

‡ Department of Geology, College of Wooster, Wooster, Ohio 44691, USA

It has been proposed that shallow carbonate seas of Upper Ordovician and Jurassic ages precipitated calcite as the primary calcium carbonate phase¹⁻³, unlike such settings today. It is expected that such seas were undersaturated with respect to the more soluble aragonite, so that aragonitic invertebrate skeletons would have tended to dissolve. Here we describe examples in which this seems to have been the case. Our evidence consists of submarine-cemented horizons in which aragonite fossils have been dissolved out to leave open moulds. The walls of the moulds are encrusted by contemporaneous organisms, confirming the early timing of the aragonite dissolution.

The current orthodoxy that shallow carbonate seas have precipitated different carbonate minerals at different times during the past 600 Myr has developed from the study of oolites¹⁻³. Modern shallow-marine oolites are predominantly aragonite and show textures (principally the lack of a radial cortical structure) that are quite different from those of most oolites from certain ancient shallow lime-precipitating seas, including most of those of Ordovician and Jurassic ages^{2,3}. These differences cannot be explained by diagenetic alteration. The ancient oolites are now generally regarded as originally having had a radial calcite cortex^{2,3}, and it has been suggested that throughout the Phanerozoic, episodes of primary inorganic precipitation of calcite have alternated with periods during which aragonite (with subordinate high-Mg calcite) has been the primary calcium carbonate precipitate. Fluctuations in Mg:Ca ratios and variations in marine CO₂ concentrations have both been suggested as possible driving causes^{1,2,4}.

The rules established for oolite formation have also been applied to other primary inorganic marine precipitates, such as the marine cements found in early diagenetic hardgrounds and cobble horizons². These ancient hard sea floors were formed by *in situ* cementation without emergence, and never show karstic, pedogenic or other emergence features. Modern analogues are invariably cemented with aragonite or high-Mg calcite, but original calcite cements appear to have been the rule in such Ordovician and Jurassic sea floors^{5,6}.

Because aragonite is more soluble than calcite at surface pressures and temperatures, seas that precipitated calcite as sea-floor cements would be expected to have been undersaturated with respect to aragonite, and aragonite of bioclastic origin would be expected to have gone into solution in the cementing

environment. The main problem in identifying this process is that aragonite typically dissolves as a result of later changes in diagenetic circumstances that are common to the history of any ancient rock, to yield a secondary biomouldic porosity that usually becomes filled by cement of later origin. It is therefore difficult to distinguish dissolution that has taken place on the sea floor from that occurring later.

Palaeoecological study of early-cemented carbonate horizons can solve this problem. Hardgrounds and early-cemented cobbles are characterized by hard-substrate-encrusting organisms, which indeed constitute one of the principal criteria for recognizing the original hard nature of such substrates in the field⁷. The association of early calcite precipitation with aragonite dissolution could have resulted in an open biomouldic porosity on the sea floor. If encrusting invertebrates colonized the walls of these pores (Fig. 1), then the early timing of aragonite dissolution can be detected. We now have clear evidence that this was a common process in at least some ancient calcite-precipitating seas. The shallow-water carbonate deposits of the north-west European Jurassic provide the best examples, and we have seen the same phenomenon in Upper Ordovician hardgrounds from Indiana and Ohio, USA.

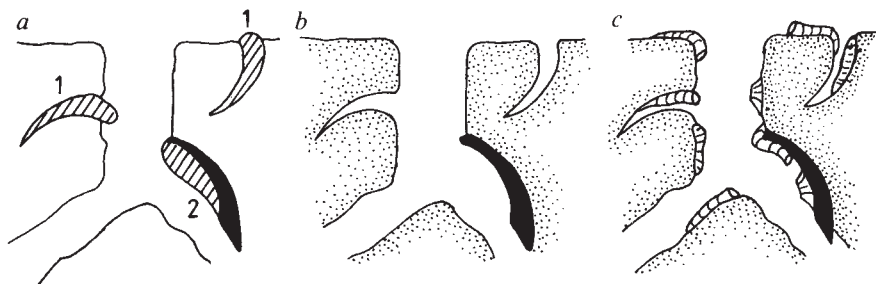
Two main stratigraphic horizons provide most of the Jurassic examples. The first is the Middle and Upper Bathonian shelf and ramp carbonate grainstones in the Cotswolds, England (White Limestone Formation⁸) and in Normandy, France (Calcaire de Ranville⁹). The second is the Upper Bathonian Lower Cornbrash, a thin bioclastic unit traceable across all of southern England and Normandy, which represents the onset of the Callovian transgression.

The stratigraphically lower examples are all associated with typical grainstone hardgrounds, some of which were penetrated by a burrowing fauna before lithification. These burrows were kept open, probably by their crustacean constructors, as lithification proceeded within the sediment around them, resulting in open burrow systems which descend from the hardground surface. Aragonite shells of molluscs originally intersected the surfaces of the hardgrounds and the burrow walls, and their early dissolution left holes (moulds). The lithified burrow walls often support a specialized fauna of gloomy-cavity hard-substrate-encrusting species (bryozoans, calcisponges, serpulid worms, thecideacean brachiopods¹⁰), and it is these forms that are seen most often encrusting the moulds that remain after aragonite dissolution (Table 1, A-D).

Some of the bivalve taxa incorporated into the soft sediment (limids and pectinids, for example) originally had shells that consisted of a calcite outer layer and an aragonite inner layer. Aragonite dissolution of such shells exposed the inner face of the outer (calcite) layer, and we have seen encrustation of these faces (Fig. 1). Borings of endolithic worms and bivalves have been excavated locally into the aragonite shell layers before shell burial and dissolution. The borings have become filled with sediment and this filling, when cemented, is revealed as a club-shaped protruberance following aragonite dissolution of the shell. We have also seen subsequent encrustation of these protruberances (Fig. 2; Table 1, B, D).

Examples of mould encrustation association with Jurassic

Fig. 1 Sequence of events taking place on certain Jurassic and Ordovician sea floors, indicating submarine dissolution of shells that were wholly composed of aragonite (hatched; 1), or that consisted of an aragonite inner shell layer and a calcite outer shell layer (black; 2). *a*, Soft sediment, containing shells, is penetrated by burrows descending from the sea bed. *b*, Sediment surface and burrow walls become lithified by precipitation of calcite cement; aragonite dissolves. *c*, sediment surface, burrow walls and moulds after aragonite dissolution, now all hardened, are encrusted by a hard-substrate-specific fauna.



hardgrounds have previously been illustrated^{11,12}, and were interpreted, in line with the orthodoxy of the time, as evidence for a period of emergence and flushing by meteoric waters. It is clearly important, therefore, to establish that the hardgrounds discussed here were not, in fact, emergent, and there is some evidence to this effect. None of the hardgrounds that we have studied shows the karstic or pedogenic features characteristic of emergence, and none is overlain by deposits with land-based or fresh-water biotas, even though such features have been seen elsewhere in the English Bathonian^{8,13}. Isotope studies of Bathonian hardgrounds taken from the Great Oolite outcrop, including the example previously interpreted as emergent¹³ (Table 1, A), have shown marine signatures and do not support the emergence interpretation¹⁴. The main evidence for emergence of certain Great Oolite hardgrounds comes from studies of sub-surface examples that cannot be matched stratigraphically with those that we have examined^{15,16}. The evidence for emergence of those examples consists of the blocky calcite nature of early cements and of the presence of over-compacted fabrics and distorted grains that are interpreted as representing early dissolution. In the light of work cited above⁶ and of the study presented here, these features do not prove emergence. Further studies of marine hardground textures should be illuminating in this respect.

Even in hardgrounds on which biological encrustation is sparse, the early timing of aragonite loss can sometimes be inferred. The residual primary porosity that remains in the rock after the development of thin isopachous submarine cement rinds on the grains is often filled by fine micrite or clay which filters down between the grains for a few centimetres below the hardground surface. Such fine sediment often overlies the submarine cement in these circumstances and should not be mistaken for vadose silt, which would indicate emergence. We have examined Bathonian hardgrounds in which such fine-grained material also fills biomouldic porosity (Table 1, A, B, E, F). We have also found examples of Bajocian hardgrounds with marine isotope signatures¹⁷ in which the biomouldic porosity itself is lined by a contribution from the early cement generation.

The examples from the Cornbrash of Normandy and England, which were developed in a transgressive episode and thus are unlikely candidates for early emergence, provide further support for the model presented here. They involve early cementation

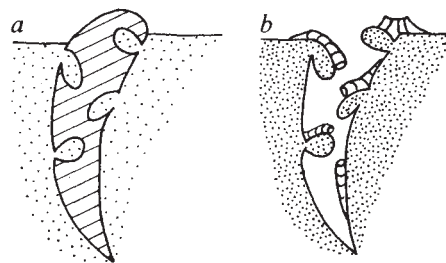


Fig. 2 Encrustation of moulds after previously-bored aragonite shells. *a*, Aragonite shell has been bored by endolithic organisms and then buried. Soft sediment surrounds the shell and fills the borings. *b*, Sediment around the shell and in the borings becomes calcite-cemented, and aragonite of the shell dissolves. Encrusting organisms colonize the mould of the shell and the lithified infills of the borings.

within microenvironments that were afforded by sediment-filled articulated bivalves (Table 1, E–G). The Lower Cornbrash of Thrapstone, Northamptonshire contains many of these cemented internal moulds, together with cemented burrow-fills, which were exhumed on the ancient sea bed to produce cobbles. These were encrusted by bryozoans, serpulids and oysters. Irregularities on the surfaces of the cemented sediment fills are moulded accurately by the attachment surfaces of the encrusters, and removal of the shell is clearly apparent. We believe that dissolution, rather than abrasion, was responsible for shell removal, because of the sharp outlines of the moulds and because we have seen the same phenomenon in Normandy on a bivalve mould that was still in its vertical life-position, and which, therefore, could never have been exhumed (Table 1, I).

Bryozoan encrustation of moulds after dissolution of aragonite cephalopods, gastropods and bivalves, is also seen in Upper Ordovician hardgrounds of the Dilsboro Formation (south-east Indiana, USA¹⁸; Table 1, J) and the Bull Fork Formation (southern Ohio, USA; Table 1, K). Again, features indicating emergence are absent. In general the phenomenon is likely to be encountered more rarely in Lower Palaeozoic rocks because of the greater rarity of groups with aragonite shells. But the Jurassic and Ordovician periods both represent times of maximum development of calcite seas² and we expect to encounter many more examples. Calcite seas seem to have been prevalent also during the Silurian, Devonian, Mississippian and Cretaceous periods², and hardgrounds of these ages should be studied for evidence of similar early aragonite dissolution. The model should be tested further through examination of hardgrounds and early-cemented lithologies developed at times when, as at present, aragonite is thought to have been the primary precipitate. Examples of early biomouldic porosity in these rocks may represent episodes of emergence and meteoric flushing, and, if so, may have a lateral extent that is sufficient to permit their use in local stratigraphic correlation.

Received 25 July; accepted 13 September 1988.

Table 1 Localities where early-cemented limestones show evidence of contemporaneous submarine dissolution of aragonite

A	Hardground in White Limestone Formation, Foss Cross Quarry, England (SP 056091).
B	Hardground at top of Pierre de Ranville, Ranville Quarry, Normandy ($x = 1173.0$, $y = 409.9$).
C, D	Hardgrounds within Caillasse de la Basse Ecarde & Caillasse de Blainville, Campagnettes Quarry, Normandy, France ($x = 1174.9$, $y = 411.5$).
E, F	Hardgrounds at top of Ardley Member and Bladon Member of White Limestone Formation, Stonelands Quarry, near Burford, England (SP 278098).
G	Cemented internal bivalve moulds, Lower Cornbrash, Shipton Quarry, near Oxford, England (SP 478171).
H	Cemented internal bivalve moulds, Lower Cornbrash, Thrapstone Quarry, England (TL 000776).
I	Cemented internal bivalve moulds, Lower Cornbrash, Ranville New Quarry, Normandy, France ($x = 1172.3$, $y = 410.4$).
J	Hardground in Dilsboro Formation, in roadcut on Station Hollow, Dilsboro, Indiana, USA, 400 m north of intersection of State Route 262 with US50.
K	Hardground in Bull Fork Formation, in roadcut on east side of Ohio State Route 136, 5.3 km north of Manchester, Ohio, USA.

A–I: Middle Jurassic (Bathonian) of England and Normandy, France (French locality references given on the Lambert zone 1 kilometre grid system); J, K: Upper Ordovician, near Cincinnati, Ohio, USA.

- Sandberg, P. A. *Sedimentology* **22**, 497–537 (1975).
- Sandberg, P. A. *Nature* **305**, 19–22 (1983).
- Wilkinson, B. H., Owen, R. M. & Carroll, A. R. *J. sedim. Petrol.* **55**, 171–183 (1985).
- Mackenzie, F. T. & Pigott, J. D. *J. geol. Soc. Lond.* **138**, 183–196 (1981).
- Wilkinson, B. H., Janekke, S. U. & Brett, C. E. *J. sedim. Petrol.* **52**, 47–57 (1982).
- Wilkinson, B. H., Smith, A. L. & Lohmann, K. C. in *Carbonate Cements* SEPM spec. Publ. 36 (eds Schneidermann, N. & Harris, P. M.) 169–184 (1984).
- Palmer, T. J. *J. Paleont.* **52**, 109–117 (1979).
- Palmer, T. J. *Palaentology* **22**, 189–228 (1979).
- Fily, G. *C.R. Acad. Sci. Paris Ser. D* **278**, 3039–3042 (1974).
- Palmer, T. J. & Fürsich, F. T. *Palaentology* **17**, 507–524 (1974).
- Fürsich, F. T. & Palmer, T. J. *Proc. geol. Ass. Lond.* **86**, 171–181 (1975).
- Fürsich, F. T. *Neues. Jb. Geol. Palaont. Abh.* **158**, 1–63 (1979).
- Palmer, T. J. & Jenkyns, H. C. *Sedimentology* **22**, 125–135 (1975).
- Marshall, J. D. thesis, Univ. Leicester (1979).
- Sellwood, B. W., Scott, J., Mikkelsen, P. & Akroyd, P. *Mar. Petrol. Geol.* **2**, 44–55 (1985).
- Sellwood, B. W., Scott, J., James, B., Evans, R. & Marshall, J. in *Petroleum Geology of North West Europe* Vol. 1 (eds Brooks, J. & Glennie, K. W.) 129–137 (Graham & Trotman, London, 1987).
- Marshall, J. D. & Ashton, M. *Sedimentology* **27**, 271–289 (1980).
- Bodenbender, B. E., Wilson, M. A. & Palmer, T. J. *Lethaia* **22** (in the press).

Carbonate dissolution in a modern mixing zone

P. L. Smart*, J. M. Dawans† & F. Whitaker*

* Department of Geography, University of Bristol, Bristol BS8 1SS, UK

† Koninklijke/Shell, Exploratie en Productie Laboratorium, Volmerlaan 6, 2280 GD Rijswijk, The Netherlands

The potential for dissolution of calcite and aragonite, and precipitation of dolomite, by the mixing of fresh and saline ground waters has been demonstrated theoretically¹⁻³. Although geologists have used these processes in the explanation of the diagenesis of recent⁴⁻⁷ and ancient⁸⁻¹² carbonates, there has been only limited investigation of the geochemical processes of modern mixing zones^{13,14}. Here we report the results of chemical and rock profiling through a modern mixing zone on Andros Island in the Bahamas. Our results demonstrate undersaturation of the waters in the mixing zone with respect to aragonite and calcite, and corresponding pervasive dissolution of the carbonate rocks at this level. Undersaturation is significantly enhanced by bacterial processes, which are also responsible for precipitation of an iron-rich crust. Despite chemical supersaturation, precipitation of dolomite was not observed, thereby casting some doubt on the mixing-zone model of dolomitization.

Beneath South Andros Island, Bahamas (Fig. 1), blue holes^{15,16,19} lead to an extensive system of underwater caves, which has been explored using specialized cave-diving techniques. In many of these caves, there is a transition from fresh or brackish waters at the surface to ground waters of seawater salinity at depth (Fig. 2), the mixing (or transition) zone typically being from 2 to 10 m in thickness.

Undisturbed groundwater samples were obtained using sample tubes filled by forward motion of a diver, and capped immediately. On return to the surface, water from the tubes was passed by a tap at the base through a stirred cell in which pH was measured using a digital pH meter calibrated with NBS standard buffers. Sub-samples were taken for immediate measurement of alkalinity and calcium by standard titration techniques, and for later analysis of other major anions and cations. In all analyses the ion-balance error was less than 5%. Saturation indices with respect to aragonite, calcite and dolomite and the carbon dioxide partial pressure (p_{CO_2}) with which the waters were in equilibrium were calculated using the WATSPEC

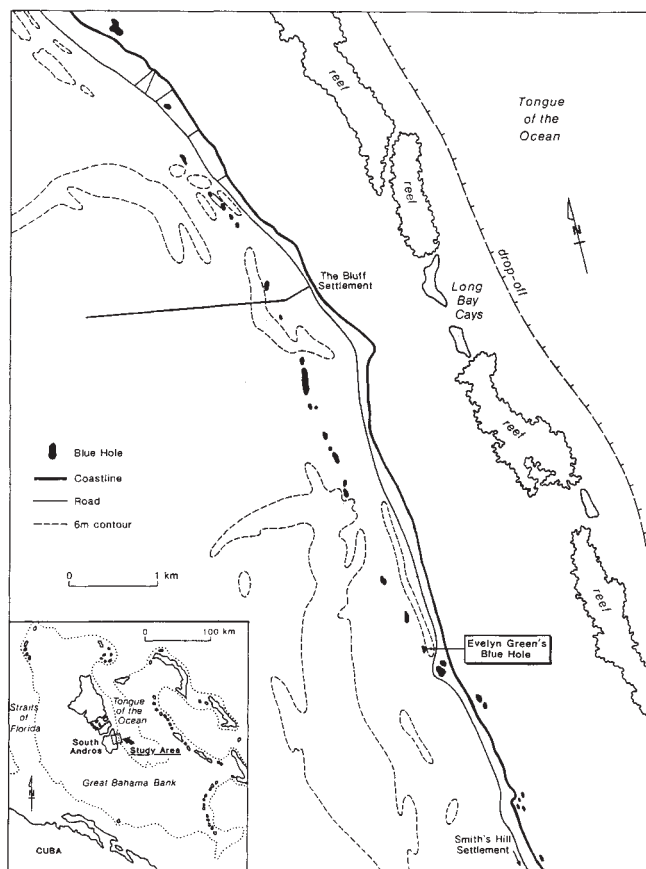


Fig. 1 Location map of the Bahamas showing study area and Evelyn Green's blue hole, one of a chain of fracture-controlled blue holes on South Andros Island.

aqueous speciation model¹⁸. In addition to the water samples, a description of the gross wall-rock morphology was made and rock samples were collected for petrographic examination. Suites of water samples were obtained from several sites and all showed similar results. Those from Evelyn Green's blue hole (Fig. 1) are representative and are described below.

The surface waters are brackish with essentially uniform salinity to 10 m depth (Fig. 2). Below this depth, salinity changes rapidly in the transition zone and reaches a value similar to that

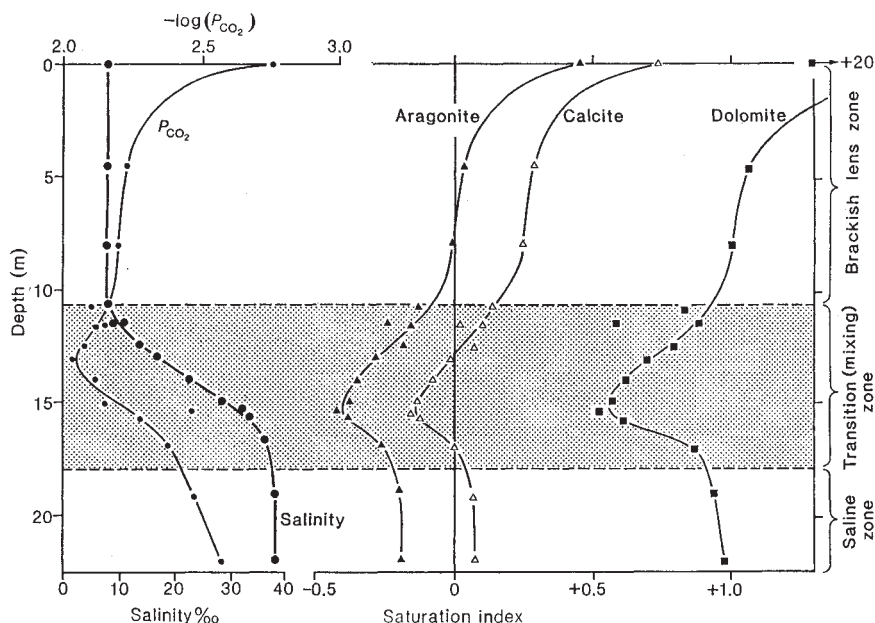


Fig. 2 Vertical variation of salinity, p_{CO_2} and aragonite, calcite, and dolomite saturation indices in Evelyn Green's blue hole.

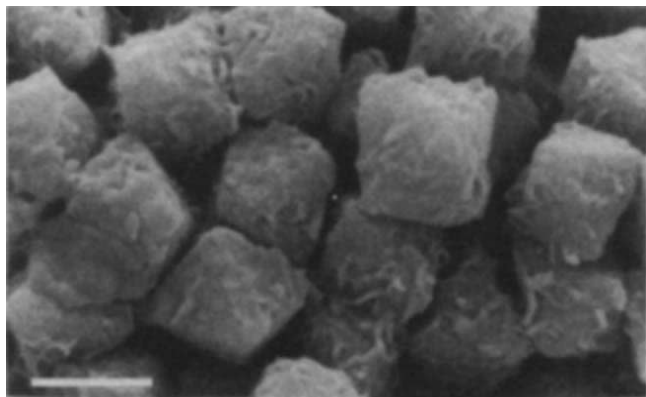


Fig. 3 SEM photomicrograph showing iron sulphide precipitated in an iron-rich crust at the base of the transition zone. Note the filaments enclosed in the crystals, possibly representing remains of sulphate-reducing bacteria. Scale bar is 1 μm .

of bank sea water at a depth of about 19 m. Constrictions prevented physical access below 22 m, the deepest sample taken. At the surface, the waters are supersaturated with respect to both calcite and aragonite, because of degassing, but are equilibrated with aragonite in the lower part of the brackish lens. Waters in the upper part of the mixing zone (below 10 m) are undersaturated with respect to aragonite, whereas below 12 m undersaturation with respect to calcite also occurred. Below 17.5 m the saline ground water was saturated or supersaturated with respect to calcite. This is the first time that waters undersaturated with respect to calcite have been reported in clear association with a mixing zone and dissolution of the adjacent bedrock might therefore be expected.

All waters sampled were supersaturated with respect to dolomite. Theoretical calculations suggest that saturation should increase through the mixing zone, as magnesium concentrations increase with salinity². This expected pattern is not evident in our results, for which dolomite saturation decreases through the mixing zone essentially in parallel with the saturation index for calcite. A possible explanation for this behaviour is that p_{CO_2} also varies through the mixing zone, obtaining a maximum value in the upper part but declining below this into the saline waters (Fig. 2). Thus, calcite and aragonite undersaturation may result from changes in p_{CO_2} . The increase in p_{CO_2} may be related to aerobic bacterial production of CO_2 by oxidation of particulate organic matter suspended in the density gradients of the mixing zone. The mixing-zone waters were highly coloured, and in the lower part of the mixing zone hydrogen sulphide was present. This suggests that the lower part of the mixing zone is anaerobic and bacterial reduction of sulphate drives organic-matter oxidation. Although precipitation of iron sulphide does occur (see below), much of the reduced sulphate is evolved in the form of gaseous hydrogen sulphide, which, on rising into oxic waters, is oxidized, generating sulphuric acid. The consequent lowered pH enhances the degree of undersaturation observed in these waters. We believe that bacterial oxidation of organic matter is an important geochemical process in modern mixing zones in the Bahamas, as has also been reported for near-surface waters¹⁹⁻²¹.

Petrographic examination of limestone samples from Evelyn Green's blue hole showed a sharp facies-transition at the top of the present mixing zone (at ~ 10 m depth) between predominance of peloidal/oolitic grainstones above and finer skeletal (algal/foram-rich) grainstone/packstone below. The gross wall-rock morphology observed in the cave was relatively smooth below the mixing zone, with increased fretting upward into the mixing zone, culminating in a very porous 'Swiss cheese' texture in the upper mixing zone. Above the facies transition and the top of the mixing zone, there was an abrupt reversion to a relatively smooth wall morphology. In the basal 2-3 m of the

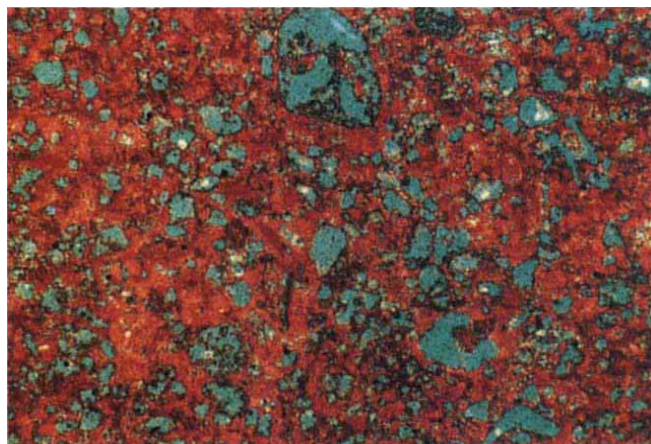
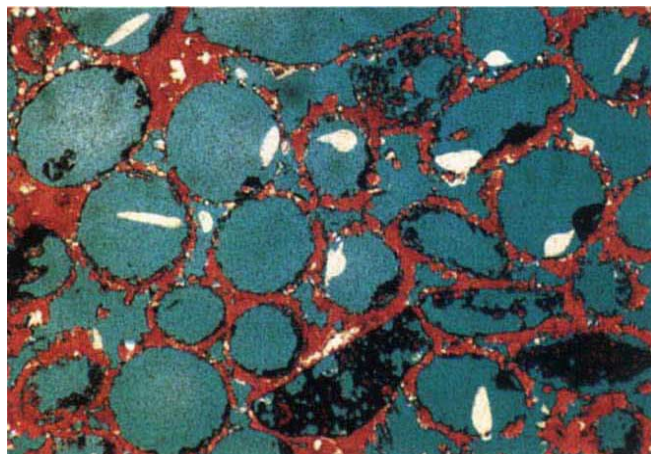
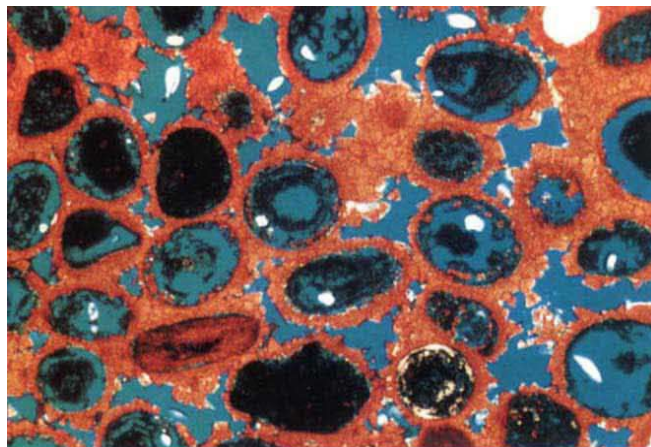


Fig. 4 Comparative photomicrographs of oolitic/peloidal limestone samples from Evelyn Green's blue hole before (top) and after (middle) mixing-zone dissolution. Field of view is 4.5×3.2 mm. Top panel, low magnesium calcite cement occludes inter-granular porosity between low-magnesium calcite ooids and peloids. Middle, both ooids and cement have been dissolved. Bottom, highly corroded, fine skeletal (foram-rich) packstone from the mixing zone. Note that degree of mixing-zone dissolution is not facies-dependent. All sections are stained with alizarin red S; porosity in the lower two panels is identified by blue plastic impregnation. Field of view 1.8×1.3 mm.

mixing zone, a black to brown crust containing precipitates of FeS_2 (composition defined by scanning electron microscope-electron-diffraction crystallographic measurements) was observed on the rock surface (Fig. 3), but no other evidence of secondary deposition was noted. Although possibly partially facies-controlled, there is a clear association between wall-rock

morphology and the present position of the mixing zone. Sea level has been around the present level for only less than 2,000–3,000 years (ref. 22), suggesting that mixing-zone dissolution may be sufficiently rapid to overprint previous diagenetic alterations within this relatively short time period. Other authors have also commented on the apparent rapidity of processes occurring in the mixing zone^{4,23}.

Within the brackish lens above the mixing zone, the extent of secondary dissolution and porosity development shows considerable variation. Some limestone samples are tightly cemented by low-magnesium calcite (Fig. 4a), but in general there is development of some oomouldic porosity by dissolution of the ooids and peloids, and of cross-cutting karstic channels, some of which are filled by microbreccia and fine sediments. At the top of the mixing zone, oomouldic porosity is well developed and there is only limited attack of the calcite cement, but below 12 m all limestone samples show a pervasive dissolution of both allochems and matrix (Fig. 4b). This correlates precisely with the observed undersaturation of the mixing-zone waters with respect to calcite, and is clearly a result of attack by these aggressive waters. The pervasive corrosion is not limited to the originally more permeable, coarser carbonates but is also developed in the relatively fine-grained carbonates such as, for example, very fine skeletal packstones (Fig. 4c). Pervasive matrix dissolution is also observed in the saline ground waters below the mixing zone. This is in contrast to the slight supersaturation of the present ground waters with respect to calcite (Fig. 2). We propose that the matrix dissolution observed in the saline zone has been inherited from a previous position of the mixing zone before sea level rose to its present position.

Although the waters are supersaturated with respect to dolomite, there is no evidence of dolomite forming in the present mixing zone, despite the occurrence of sulphate reduction²⁴. A possible explanation for the lack of dolomite is that dolomitization may be a much slower process than calcite dissolution in this environment²⁵, as suggested by Plummer².

The strong circulation which occurs in these cavern systems beneath South Andros Island is particularly favourable for groundwater mixing; this aspect will be discussed elsewhere. Our study also suggests, however, that the density gradients present in the mixing zone cause accumulation of particulate organic matter at this level, and thus bacterial processes may also play a significant role in driving carbonate dissolution. We are currently investigating the bacterial processes that occur and the extent to which these processes are significant elsewhere in driving carbonate dissolution. The apparent absence of dolomite in the modern mixing zone beneath South Andros Island suggests serious limitations in the explanation of the thick sequence of late Cenozoic dolomites beneath the Bahamian platform²⁶ by the 'Dorag' or mixing zone dolomitization model^{6,8,27}.

We thank the Royal Society and particularly Shell UK Exploration and Production for financial support, and Shell Research BV and Shell International Petroleum Maatschappij for permission to publish this paper. Logistic support from British Airways Cargo, Virgin Atlantic, Shell Bahamas, Neil Sealey, Joy Chaplin and Commander I. Anderson (Royal Navy) is also acknowledged. Finally we thank the Government of the Bahamas for permission to undertake the study, and particularly Richard Cant of Public Works and Utilities, and the Lands and Forests Departments for their help.

Received 20 June; accepted 19 August 1988.

1. Bogli, A. *Erdkunde* **18**, 83–92 (1964).
2. Plummer, L. N. *Geol. Soc. Am. Mem.* **142**, 219–236 (1975).
3. Wigley, T. M. L. & Plummer, L. N. *Geochim. cosmochim. Acta* **40**, 989–995 (1976).
4. Humphrey, J. D. thesis, Brown Univ. (1987).
5. Land, L. S. & Epstein, S. *Sedimentol.* **14**, 187–200 (1970).
6. Ward, W. C. & Halley, R. B. *J. sedim. Petrology* **55**, 407–421 (1985).
7. Magaritz, M., Goldenberg, L., Kafri, U. & Arad, A. *Nature* **287**, 622–624 (1980).
8. Badiazamani, K. *J. sedim. Petrology* **43**, 965–984 (1973).
9. Choquette, P. W. & Steinen, R. P. *Soc. econ. paleontol. Miner. spec. Publ.* **28**, 163–196 (1980).
10. Dunham, J. B. & Olson, E. R. *Geology* **6**, 556–559 (1978).

11. Sears, S. O. & Lucia, F. J. *Soc. econ. paleontol. Miner. spec. Publ.* **28**, 215–235 (1980).
12. Meyers, W. J. & Lohman, K. C. *J. sedim. Petrology* **48**, 475–488 (1978).
13. Back, W., Hanshaw, B. B., Pyle, T. E., Plummer, L. N. & Weidie, A. E. *Wat. Resour. Res.* **15**, 1521–1535 (1979).
14. Randazzo, A. F. & Cook, D. J. *sedim. Geol.* **54**, 169–192 (1987).
15. Smart, P. L., Palmer, R. J., Whitaker, F. & Wright, V. P. in *Paleokarst* (eds James, N. P. & Choquette, P. W.) 149–163 (Springer, New York, 1987).
16. Dill, R. F. *Proc. 3rd int. Coral Reef Symp* 237–242 (1977).
17. Palmer, R. *New Scientist* **1507**, 44–48 (1986).
18. Wigley, T. M. L. *Br. Geomorph. Res. Grp Tech. Bull.* **20** (Norwich, 1977).
19. Geblein, C. D. *et al. Soc. econ. paleontol. Miner. spec. Publ.* **28**, 31–49 (1980).
20. Pigott, J. D. & Land, L. S. *Mar. Chem.* **19**, 355–378 (1986).
21. Thorstenson, D. C. & Mackenzie, F. T. *Geochim. cosmochim. Acta* **38**, 1–19 (1974).
22. Wanless, H. R. *J. sedim. Petrology* **52**, 1051–1054 (1982).
23. Back, W., Hanshaw, B. B., Herman, J. S. & van Driel, J. N. *Geology* **14**, 137–140 (1986).
24. Baker, P. A. & Kastner, M. *Science* **213**, 214–216 (1981).
25. Walter, L. M. *Soc. econ. paleontol. Miner. spec. Publ.* **38**, 1–12 (1986).
26. Dawans, J. M. & Swart, P. K. *Sedimentol.* **35**, 385–403 (1988).
27. Hardie, L. A. *J. sedim. Petrology* **57**, 166–183 (1987).

The costs of reproduction in the collared flycatcher *Ficedula albicollis*

Lars Gustafsson & William J. Sutherland

Department of Zoology, Box 561, 5751 22 Uppsala, Sweden and School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK

There has been considerable interest in testing Lack's¹ hypothesis that the most frequent clutch size should be the one which yields the highest number of viable offspring. Many tests, however, have shown that the clutch size that produces the most fledglings is actually larger than the most frequent size^{2,3}. This discrepancy has been accounted for by suggesting that high fecundity may result in a reduction in fledgling survival¹, juvenile survival⁴, adult survival^{5,6}, subsequent fecundity⁷ or the fecundity of the offspring⁸ or, alternatively, that there is a relationship between parent quality or territory quality and brood size⁹. We evaluate three different methods of determining these costs and conclude that experimental manipulations provide the only reliable means of determining the costs. The collared flycatcher, *Ficedula albicollis*, on the island of Gotland forms an isolated population with a remarkably high site fidelity of both adults and young such that the lifetime reproductive success can be assessed with an accuracy rarely possible under natural conditions. Experimental modifications of the actual clutch size showed that the main costs of an enlarged clutch were that the juveniles were less likely to survive, that the subsequent fecundity of the parents was reduced and the offspring had a lower fecundity. Adult survival was unaffected by clutch size.

The clutch size that results in the highest fitness is a compromise between the numerical benefits of producing additional young and the costs associated with these young^{1,10}. Three methods have been used to analyse these costs. Relationships between clutch size and the other life history characters have been investigated by comparing species with different clutch sizes^{11,12}, by comparing individuals within a population^{1,4,13}, or by experimentally manipulating clutch size^{3,5,6}. Using the first technique, O'Connor¹² showed that European bird species with high fecundity had low survival. However there is an inevitable relationship between demographic parameters for any approximately stable population^{14,15}. Assuming no long-term increase or decrease in population size it is possible to calculate the relationship between demographic parameters for a species which first breeds when n years old. The relationship between F , the average fecundity of each female over n years old, S_a , the adult survival, and S_i , the values of survival for each of the i years before breeding is

$$F = \frac{2(1 - S_a)}{\sum_{i=1}^n S_i} \quad (1)$$

The relationship between fecundity, age at first breeding and

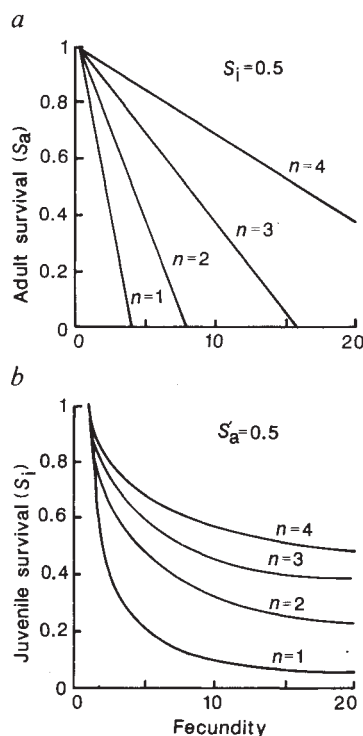


Fig. 1 Relationship between fecundity and (a) adult survival and (b) juvenile survival as determined by equation (1).

both juvenile and adult survival is shown in Fig. 1. For a given age at first breeding, higher fecundity results in lower survival of either adults, juveniles or both. Showing a simple relationship between such variables cannot be considered as evidence for a life history trade-off.

The important issue in comparing species is not whether any relationships exist, as some relationships are inevitable, but rather, which of the possible combinations occur. Thus a species or population with high adult mortality may be balanced by either high fecundity, low juvenile mortality, low age of first breeding, or any combination of these. Sæther's analysis¹¹ suggests that all three parameters are correlated with survival. Precisely which combinations occur will depend upon the interaction of evolutionary selection pressures and the precise way in which density dependence operates^{11,16}, but this approach can never reveal the adaptive significance of life-history traits.

The pied flycatcher *Ficedula hypoleuca* is one of the European species included in Sæther's¹¹ comparative analysis. We have applied the other techniques for measuring life-history trade-offs to its allospecies, the collared flycatcher, *Ficedula albicollis*, by observing and manipulating the clutches of individuals within a population. The species was studied from 1980 to 1987 on the island of Gotland in the Baltic.

Parent birds were caught, with traps in the nest boxes, while feeding the young and marked with unique numbered rings. The numbers of fledglings in each brood were counted 13 days after hatching and each was uniquely ringed, weighed, and its tarsus measured. Surviving adults and the number of offspring recruited into the breeding site were determined by catching all breeding pairs within the study area in subsequent years. There is a remarkably high rate of return of adults and young to the island¹⁷⁻¹⁹. The breeding success of these recruits was also measured.

During three years (1983-5) the clutches of 320 collared flycatchers were manipulated. Clutches with the same hatching date were treated in pairs and either one or two young, at two days of age, were moved from one nest to the other, or for the control group, two young were swapped between nests. Previous

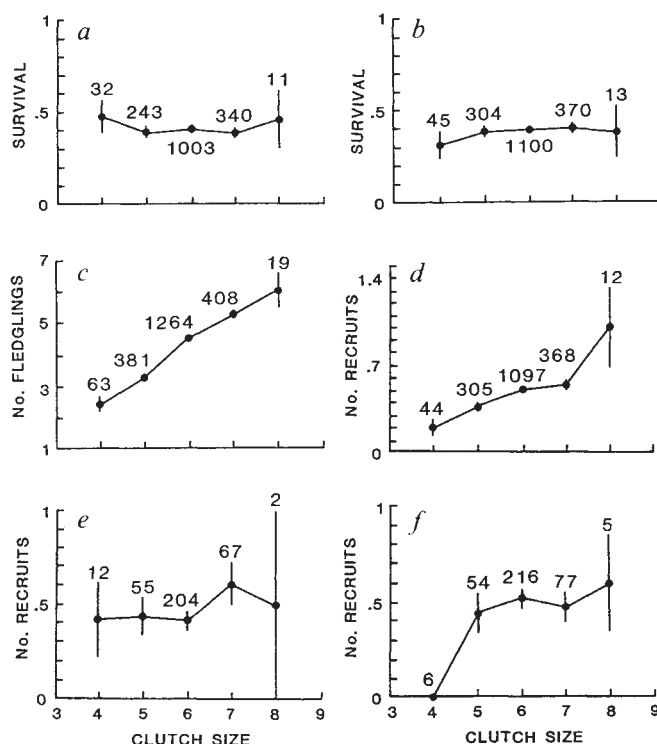


Fig. 2 Relationship between the unmanipulated clutch size and (a) adult male survival (regression: $y = 0.43 + 0.01x$, $n = 1,629$, $P > 0.6$), (b) adult female survival (regression: $y = 0.34 + 0.01x$, $n = 1,832$, $P > 0.4$), (c) number of fledglings (regression: $y = 0.53 + 1.0x$, $n = 2,135$, $P < 0.0001$), (d) number of recruits to the breeding population (regression: $y = 0.07 + 0.11x$, $n = 1,826$, $P < 0.0001$), (e) fecundity of females in the subsequent year expressed as number of recruits (regression: $y = 0.03 + 0.07x$, $n = 344$, $P > 0.1$), (f) fecundity of female offspring expressed as number of recruits (regression: $y = 0.19 + 0.05x$, $n = 359$, $P > 0.3$). All data ± 1 standard error.

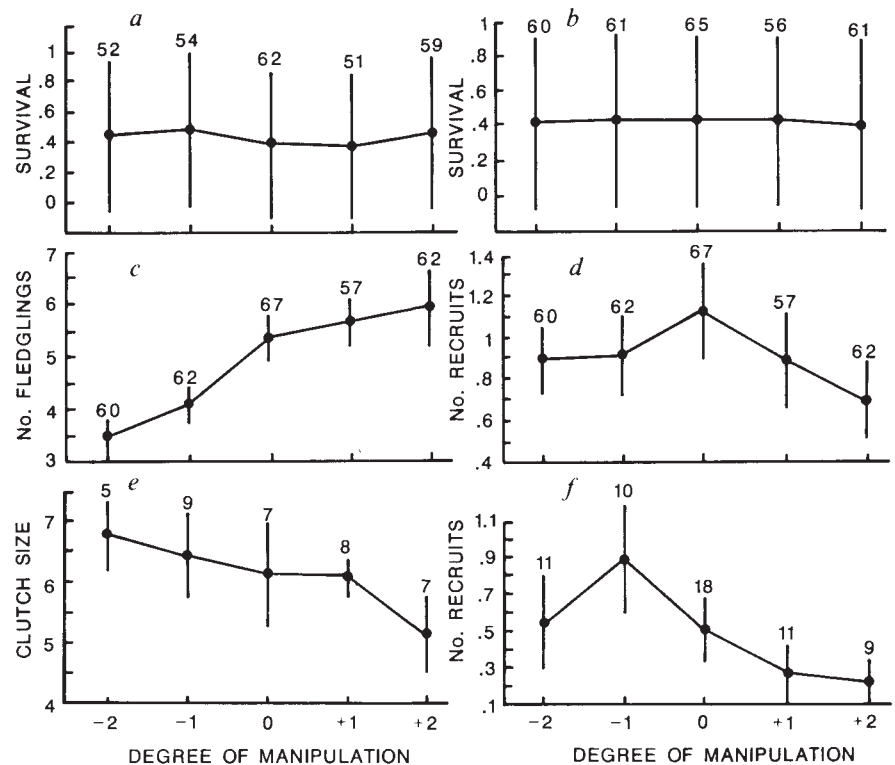
studies^{5,20} on the costs of reproduction have randomly allocated clutches of between 3 and 15 eggs with the result that an individual's clutch might be increased or decreased with up to 12 eggs. The experimental design used here, in which the natural clutch size was enlarged or reduced by one or two, has the advantage that it is nearer to the expected heritable variation.

The relationships between natural clutch size and six different parameters are shown in Fig. 2. There is no evidence, from this analysis, that clutch size had any effect on adult survival, juvenile survival or future reproduction of either the parents or young. As a consequence, larger clutches produced more recruits. This result is similar to that of many other studies of natural clutch size^{2,4,13,20,22}.

The effects of manipulating clutch size on life history parameters are shown in Fig. 3. The survival of both males and females was unaffected by the manipulation, unlike glaucous-winged gulls *Larus glaucescens*⁶ and female blue tits *Parus caeruleus*⁵.

Consistent with our observations on natural clutches, artificially enlarged clutches produced more fledged young (Fig. 3). However these larger clutches resulted in lighter and smaller nestlings (unpublished data) which were less likely to return as recruits. The result was that the natural clutch size produced more recruits than an enlarged or reduced clutch. Both the number of second generation recruits and the subsequent fecundity of manipulated females was affected by brood-size manipulation. It has been suggested^{21,22} that the trade-offs, and thus the optimal clutch size, may vary between years because of vagaries in weather and food supply. As a consequence it is unwise to draw conclusions from the results of a single year. Figure 3 represents the pooled data for three successive years,

Fig. 3 Relationship between the degree of manipulation of brood size and (a) adult male survival (regression: $y = 0.45 - 0.007x$, $n = 288$, $P > 0.07$), (b) adult female survival (regression: $y = 0.43 - 0.005x$, $n = 303$, $P > 0.8$), (c) number of fledglings (regression: $y = 4.91 + 0.65x$, $n = 308$, $P < 0.001$), (d) number of recruits (polynomial regression; $y = 1.029 - 0.044x - 0.664x^2$, $n = 308$, $P < 0.02$), (e) fecundity of experimental females in the subsequent year expressed as number of eggs (regression: $y = 7.22 - 0.36x$, $n = 36$, $P < 0.001$). (f) Fecundity of female offspring expressed as number of recruits (regression: $y = 0.475 - 0.145x$, $n = 59$, $P < 0.05$). All data ± 1 standard error.



but the pattern was consistent also when each year was considered separately. In conclusion, the major cost of an increased clutch size are the reduced recruitment of offspring, the lowered fecundity of these offspring and the reduced future offspring of females.

The difficulties in determining life-history trade-offs are well illustrated here. For demographic parameters comparisons between species or populations do not measure the costs of reproduction. There is no need for the pattern between species to correspond with the trade-offs acting upon an individual. Within a population, correlations between observed life-history parameters may fail to reveal the nature or strength of the trade-offs, presumably because the specific form of the trade-off differs between phenotypes and individuals adjust their clutch size accordingly^{9,23}. Experimental manipulations provide the only reliable means of determining the costs of reproduction.

We thank G. Carlen, M. Linden, B. Malmberg and T. Pärt for help with the field work, and P. H. Harvey, R. Nichols, C. Townsend, S. Ulfstrand and A. R. Watkinson for comments on the manuscript. This study was supported by a grant from the Swedish National Research Council to L.G.

Received 6 July; accepted 9 September 1988.

1. Lack, D. *The Natural Regulation of Animal Numbers* (Clarendon, Oxford, 1954).
2. Lessells, C. M. J. *Anim. Ecol.* **55**, 669-689 (1986).
3. Murphy, E. C. & Haukioja, E. *Current Ornith.* **4**, 141-180 (1986).
4. Perrins, C. M. & Moss, D. J. *Anim. Ecol.* **44**, 695-706 (1975).
5. Nur, N. J. *Anim. Ecol.* **53**, 479-496 (1984).
6. Reid, M. V. *Oecologia* **74**, 458-467 (1987).
7. Williams, G. C. *Am. Nat.* **100**, 687-690 (1976).
8. Anderson, M. *Am. Nat.* **112**, 762-766 (1978).
9. Högstedt, G. *Science* **210**, 1148-1150 (1980).
10. Charnov, E. L. & Krebs, J. R. *Ibis* **116**, 217-219 (1974).
11. Sæther, B.-E. *Nature* **331**, 616-617 (1988).
12. O'Connor, R. J. in *Behavioural Ecology* (eds R. M. Sibly & R. H. Smith) 105-142 (Blackwell, Oxford 1985).
13. Smith, J. M. N. *Evolution* **35**, 1142-1148 (1981).
14. Sutherland W. J., Grafen, A. & Harvey, P. H. *Nature* **320**, 88 (1986).
15. Sibly, R. M. & Callow, P. (eds) *Physiological Ecology of Animals* (Blackwell, Oxford, 1986).
16. Bennett, P. M. & Harvey, P. H. *Nature* **333**, 216 (1988).
17. Gustafsson, L. *Am. Nat.* **128**, 761-764 (1986).
18. Gustafsson, L. *Ecology* **68**, 291-296 (1987).
19. Pärt, T. & Gustafsson, L. *J. Anim. Ecol.* (in the press).
20. Nur, N. *Evolution* **42**, 352-362 (1988).
21. van Noordwijk, A. J., van Balen, J. H. & Scharloo, W. *Neth. J. Zool.* **31**, 342-372 (1981).
22. Boyce, M. S. & Perrins, C. M. *Ecology* **68**, 142-153 (1987).
23. Bell, G. *Evolution* **38**, 300-313 (1987).

Inhibition contributes to orientation selectivity in visual cortex of cat

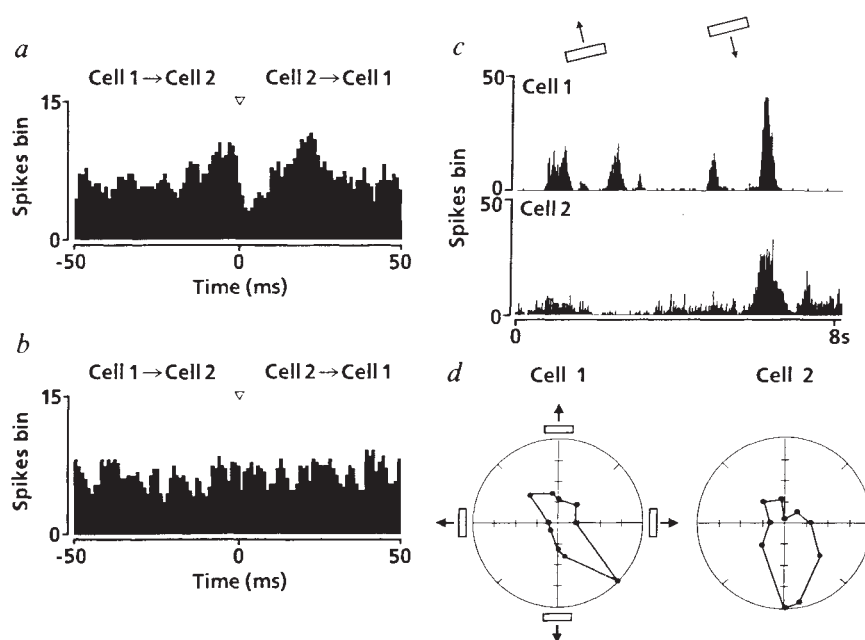
Y. Hata, T. Tsumoto, H. Sato, K. Hagihara & H. Tamura

Department of Neurophysiology, Biomedical Research Center, Osaka University Medical School, Kitaku, Osaka 530, Japan

Neurons in the visual cortex are selectively responsive to light or dark bars presented at particular orientations¹. On the basis of physiological data, this orientation selectivity is hypothesized as being due at least partially to intracortical inhibitory mechanisms²⁻⁶. But this hypothesis has been challenged by intracellular recordings indicating that excitatory inputs themselves are orientation-selective, so inhibition may not contribute to the observed selectivity⁷. Also, there is controversy about the presence of intracortical horizontal connections mediating inhibition for selectivity⁸⁻¹¹ and about the theoretical validity of such inhibitory connections¹²⁻¹⁴. Using cross-correlation analysis of the activities of two neurons recorded simultaneously, we find that inhibitory interactions exist between cells with somewhat different, but not orthogonal, orientation preferences. This suggests that intracortical horizontal inhibition operates between 'orientation columns' to sharpen the orientation tuning of cortical neurons.

Experimental procedures for preparing and maintaining cats, visual stimulation and extracellular single unit recordings have been described previously¹⁵. During recordings the animals were anaesthetized with a gas mixture of 70% N₂O:30% O₂, and paralysed with gallamine-triethiodide (Flaxedil). The heart rate, rectal temperature and expired CO₂ were monitored continuously. If the heart rate was changed suddenly by noxious stimuli, 0.1-0.5% halothane was added to the gas to maintain an adequate depth of anaesthesia. Spike discharges of pairs of cells were recorded simultaneously from primary visual cortex with two glass micropipettes which could be advanced indepen-

Fig. 1 Examples of recordings from a pair of cells showing inhibitory interaction. *a*, Cross-correlogram, representing the number of spikes per bin that occurred in cell 1 before (left half of the correlogram) and after (right half) spikes in cell 2. *b*, Shuffled cross-correlogram, calculated from spike trains in which one spike train is shifted in time by one visual stimulus period^{9,19,21}. Shuffled cross-correlograms provide information on whether the peaks and troughs of the cross-correlogram are likely to be due to stimulus coordination artefact. Auto-correlograms of the activities of both neurons which are not shown here confirmed that there was no periodicity or refractoriness of firing that might be reflected in the cross-correlogram¹⁹. In *a* and *b*, to uncover the features hidden by the high noise level, correlograms were averaged over three adjoining bins. Number of spikes: cell 1, 1,442; cell 2, 1,864. Bin width, 1 ms. *c*, Peristimulus time histograms of each cell of the pair to a light slit moving back and forth over the receptive field. Direction of motion and orientation of the slit are indicated at the top. 150 sweeps for each histogram. Bin width, 16 ms. *d*, Orientation tuning diagrams for each cell. On the basis of the cumulative number of spikes in the response profiles of the peristimulus time histograms, as shown in *c*, responses were normalized and plotted as polar diagrams. Direction of motion and orientation of the slit are indicated around the diagram for cell 1.



dently. The pipettes were filled with 0.5 M sodium acetate containing 2% Pontamine Sky Blue for histological identification of recording sites. The present analysis focused on horizontal connections, so care was taken to avoid locating the tips of both electrodes above each other, normal to the cortical surface. In some experiments multibarrel micropipettes were attached to the recording pipette to allow pharmacological activation of cortical cells. The receptive-field position and properties of each unit were explored initially with light slits produced by a handheld projector. For the quantitative analysis of receptive field properties, peristimulus time histograms of unitary spikes to slits moving at the optimal orientation were constructed with an electronically controlled projector system and by computer. The orientation of the slit was then changed in steps of 5–45° so that a polar plot of responses could be obtained. From the pair of spike trains, we constructed cross-correlograms, shuffled cross-correlograms and auto-correlograms by computer, which were displayed on-line with a monitor oscilloscope during and following their calculation. When the rate of discharge of cortical neurons was very low, a small amount of glutamate was ejected ionophoretically from the drug pipettes to sample a sufficient number of spikes. Care was taken not to activate cells with rates of discharge higher than 10 Hz.

This report considers 10 pairs of cells that showed significant inhibitory interactions in the cross-correlograms. These 10 pairs were found among 310 cell pairs that were separated horizontally by less than 1 mm, which is the approximate diameter of a 'hypercolumn'¹⁶ that includes all the orientations in the cat's visual cortex^{17,18}. Other types of interaction will be described elsewhere.

An example of the inhibitory interactions observed is shown in Fig. 1. The cross-correlogram showed a large asymmetric trough in the direction of cell 2 to cell 1 (Fig. 1*a*). The onset latency of this trough was 1 ms, and its duration was 10 ms. The shape and time course of the trough resembled those of inhibitory postsynaptic potentials (i.p.s.ps). Thus, the trough could be interpreted as a direct inhibitory synaptic connection from cell 2 to cell 1 (ref. 19). The peak preceding the trough was probably due to the presynaptic dead-time, during which there is less-than-average inhibition and therefore a higher-than-average rate of firing, and that following the trough was due to

rebound activity following i.p.s.ps of the postsynaptic cell²⁰. The shuffled cross-correlogram was quite flat (Fig. 1*b*), indicating that the visual stimuli made no significant contribution to the trough shown in the cross-correlogram. Both of the cells of this pair were located in layer IV of the cortex, and were separated horizontally by about 700 μ m. The receptive fields of these two cells overlapped each other almost completely, but their optimal orientations were slightly different, so that motion of a slit presented at an orientation intermediate between the two preferred ones elicited brisk visual responses in either cell (Fig. 1*c*). The orientation tuning diagrams of the visual responses of this pair indicate that their optimal orientations differed by 45° (Fig. 1*d*). The average onset latencies and durations of the troughs in the cross-correlograms for the eight pairs were 1.4 ± 0.7 (s.d.) and 18.6 ± 12.5 ms respectively. The other two pairs had exceptionally long onset latencies (8 and 10 ms).

To investigate the possible relationship between inhibitory connections and orientation selectivity, we established a classification scheme with four categories on the basis of differences in optimal orientation (Fig. 2). Six of the ten cell pairs showing inhibitory interactions fell into grade 2, and the

Table 1 Laminar distribution of cell pairs showing inhibitory interactions

Combination of layers	Number of pairs
Same layers II/III	0 (33)
IV	5 (82)
V	2 (38)
VI	2 (48)
Neighbouring layers	1*(100)
Others	0 (9)

Cell pairs were divided according to the location of each cell in the cortical layers. 'Others' indicates pairs consisting of cells in layers separated by more than one layer. In parentheses are given the number of pairs studied in each group. For identification of the intracortical location of cells recorded, extracellular dye marks were made through the recording pipettes and the cortex was processed later for histology.

* We found inhibitory effects in one pair from a cell in layer VI to a cell in layer V.

other four pairs into grade 1. None of the pairs with differences in optimal orientation of more than 46° (grades 3 and 4) manifested any inhibitory interaction. That is, inhibitory interactions were found mostly in the cell pairs with slightly different or rather similar orientation preferences. There were no notable differences in the magnitude of the inhibition between the cell pairs of grade 2 and those of grade 1. The mean horizontal distances between the pairs of cells in grades 1 and 2 were 231 ± 125 and $362 \pm 241 \mu\text{m}$ respectively. As to the laminar location of the cells, nine pairs with inhibitory interactions were in the same layer and the remaining one was in a neighbouring layer (Table 1). Half of the pairs having inhibitory interactions were located in cortical layer IV.

The present results demonstrate that horizontal inhibitory interactions exist between visual cortical neurons having somewhat different orientation preferences, but appear not to exist between those displaying orthogonal preferences. This conclusion provides evidence differing from previous conflicting results⁸⁻¹¹. The negative conclusion of Ts'o *et al.*⁹ on the existence of horizontal inhibitory interactions may be due to their restriction of sampling to cells from layers II/III of the cortex. In these layers, their samplings may have been biased strongly towards recordings from pyramidal cells, which are believed to be connected by excitatory synapses^{10,11}. In fact, previous studies using the cross-correlation technique have been able to detect inhibitory interactions in visual cortex, although they have been found mainly in a vertical (that is, intracolumnar) direction^{21,22}. It has been reported, however, that cross-correlation analysis is a technique that can be relatively insensitive to inhibitory interactions²³. Accordingly, we cannot completely exclude a possibility that we may have underestimated the extent of the existence of horizontal inhibitory interactions. In this sense, our findings are not altogether inconsistent with the suggestion of Matsubara *et al.*^{8,24} that patchy horizontal connections exist between cells with orthogonal orientation preferences and that these connections may inhibit each other. It should be pointed out, however, that the functional effectiveness of such connections may be insufficiently strong to be detectable by the present methods, even if they exist. So, it seems reasonable to propose that cross-orientation inhibition effectively operates for cell-cell interactions between groups displaying similar orientation preferences to quite dissimilar but not orthogonally opposed preferences.

This conclusion is consistent with anatomical evidence. Several types of visual cortical neurons that are thought to be inhibitory have axons extending horizontally by about 1 mm in length, in addition to having major vertical projections²⁵. For example, a type of small basket cell called the 'clutch cell' has horizontal axonal projections extending 300–500 μm in layer IV (ref. 26), in which most horizontal inhibitory interactions have been found in the present study. Also, the observation that horizontal projections of axons of inhibitory neurons are limited to the same layer or to neighbouring layers²⁵ is consistent with the present findings.

These results support data from previous psychophysical experiments suggesting that there is a mutual inhibition between orientation detectors with slightly different preferences²⁷, and are consistent with physiological experiments indicating that inhibitory interactions are tuned broadly to the same orientation as excitatory inputs so as to sharpen the orientation tuning of cortical cells^{3,28,29}. This conclusion seems at odds with the intracellular records of Ferster⁷ indicating that i.p.s.ps of cortical cells are tuned to orientations as sharply as excitatory postsynaptic potentials. There is the possibility, however, that inhibitory synapses may be distantly located from the soma from which his recordings were made, or additionally, a 'shunting inhibition' without a sizable hyperpolarization might have a role in selectivity. Although there are data arguing against this latter possibility³⁰, presynaptic inhibition might also be involved in helping to make cortical cells selective in their responsiveness.

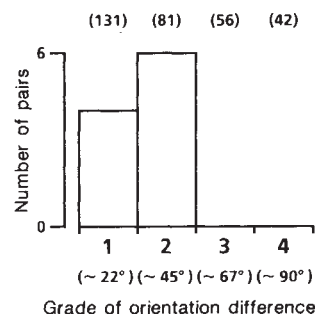


Fig. 2 Histogram displaying the number of pairs with inhibitory interaction falling in each of four optimal orientation-difference groups. Each group represents pairs of cells having orientation differences: 1, between $0-22^\circ$; 2, $23-45^\circ$; 3, $46-67^\circ$; 4, $68-90^\circ$. Total number of pairs sampled in each group is indicated at the top.

We thank Dr T. Philip Hicks for improving the English of the manuscript. This work was supported by a Grant-in-Aid for Special Project Research of Plasticity of Neural Circuits from the Japanese Ministry of Education, Science and Culture (T.T.).

Received 15 August; accepted 23 September 1988.

- Hubel, D. H. & Wiesel, T. N. *J. Physiol., Lond.* **160**, 106–154 (1962).
- Bishop, P. O., Coombs, J. S. & Henry, G. H. *J. Physiol., Lond.* **219**, 659–687 (1971).
- Blakemore, C. & Tobin, E. A. *Expl Brain Res.* **15**, 439–440 (1972).
- Benevento, L. A., Creutzfeldt, O. D. & Kuhn, V. *Nature* **238**, 124–126 (1972).
- Sillito, A. M. *J. Physiol., Lond.* **250**, 305–329 (1975).
- Tsumoto, T., Eckart, W. & Creutzfeldt, O. D. *Expl Brain Res.* **34**, 351–363 (1979).
- Ferster, D. *J. Neurosci.* **6**, 1284–1301 (1986).
- Matsubara, J., Cynader, M., Swindale, N. V. & Stryker, M. P. *Proc. natn. Acad. Sci. U.S.A.* **82**, 935–939 (1985).
- Ts'o, D. Y., Gilbert, C. D. & Wiesel, T. N. *J. Neurosci.* **6**, 1160–1170 (1986).
- Kisvarday, Z. F. *et al. Expl Brain Res.* **64**, 541–552 (1986).
- LeVay, S. *J. comp. Neurol.* **269**, 265–274 (1988).
- Rose, D. *Vision Res.* **19**, 533–544 (1979).
- Koch, C. & Poggio, T. in *Models of the Visual Cortex* (eds Rose, D. & Dobson, V. G.) 408–419 (Wiley, New York, 1985).
- Koch, C. *Soc. Neurosci. Abstr.* **13**, 1451 (1987).
- Tsumoto, T., Masui, H. & Sato, H. *J. Neurophysiol.* **55**, 469–483 (1986).
- Hubel, D. H. & Wiesel, T. N. *J. comp. Neurol.* **158**, 267–294 (1974).
- Albus, K. *Expl Brain Res.* **24**, 181–202 (1975).
- Schoppmann, A. & Stryker, M. P. *Nature* **293**, 574–576 (1981).
- Moore, G. P., Segundo, J. P., Perkel, D. H. & Levitan, H. *Biophys. J.* **10**, 876–900 (1970).
- Bryant Jr, H. L., Marcos, A. R. & Segundo, J. P. *J. Neurophysiol.* **36**, 205–225 (1973).
- Toyama, K., Kimura, M. & Tanaka, K. *J. Neurophysiol.* **46**, 202–214 (1981).
- Michalski, A., Gerstein, G. L., Czarkowska, J. & Tarnecki, R. *Expl Brain Res.* **51**, 97–107 (1983).
- Aertsen, A. H. J. & Gerstein, G. L. *J. Brain Res.* **340**, 341–354 (1985).
- Matsubara, J. A., Cynader, M. S. & Swindale, N. V. *J. Neurosci.* **7**, 1428–1446 (1987).
- Somogyi, P., Kisvarday, Z. F., Martin, K. A. C. & Whitteridge, D. *Neuroscience* **10**, 261–294 (1983).
- Kisvarday, Z. F., Martin, K. A. C., Whitteridge, D. & Somogyi, P. *J. comp. Neurol.* **241**, 111–137 (1985).
- Blakemore, C., Carpenter, R. H. S. & Georgeson, M. A. *Nature* **228**, 37–39 (1970).
- Morrone, M. C., Burr, D. C. & Maffei, L. *Proc. R. Soc. B* **216**, 335–354 (1982).
- Ramo, A. S., Shadlen, M., Skottun, B. C. & Freeman, R. D. *Nature* **321**, 237–239 (1986).
- Douglas, R. J., Martin, K. A. C. & Whitteridge, D. *Nature* **332**, 642–644 (1988).

Neuronal correlate of visual associative long-term memory in the primate temporal cortex

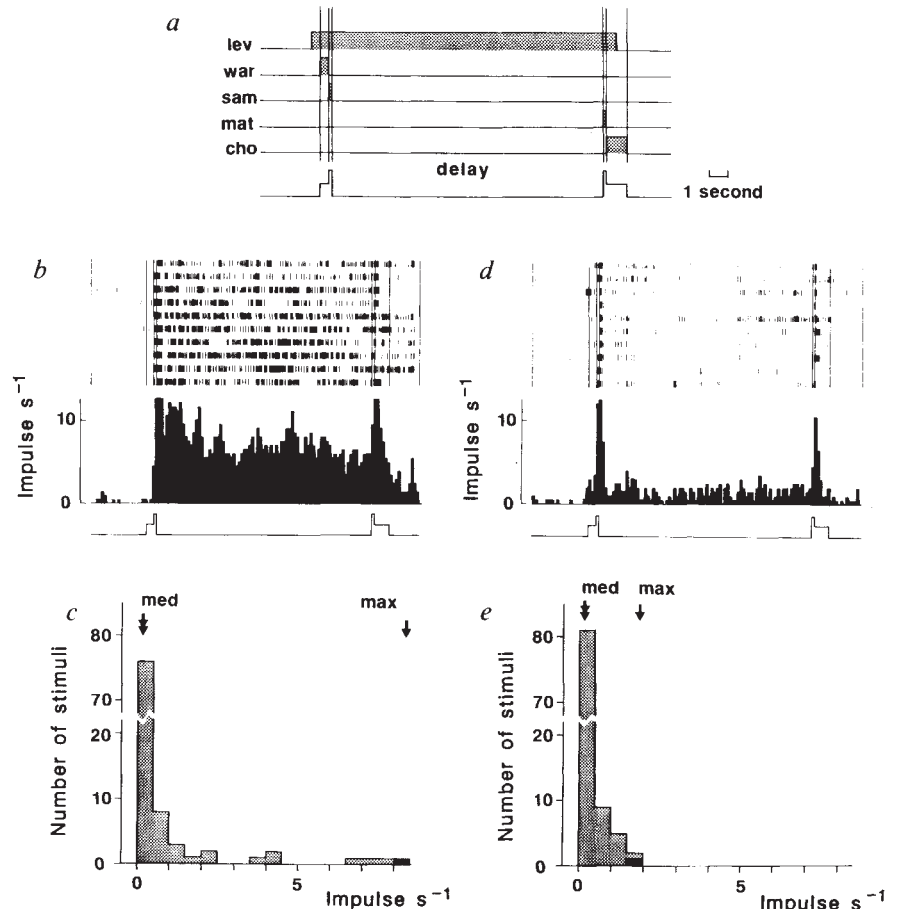
Yasushi Miyashita

Department of Physiology, Faculty of Medicine,
University of Tokyo, 7-3-1 Hongo, Bunkyo-ku,
Tokyo 113, Japan

In human long-term memory, ideas and concepts become associated in the learning process¹. No neuronal correlate for this cognitive function has so far been described, except that memory traces are thought to be localized in the cerebral cortex; the temporal lobe has been assigned as the site for visual experience because electric stimulation of this area results in imagery recall² and lesions

Fig. 1 Neuronal activity in the anterior ventral temporal cortex to the 'learned' and 'new' stimuli in a visual memory task. *a*, Sequence of events in a trial. Lev, lever press by the monkey; war, warning green image; sam, sample stimulus; mat, match stimulus following a 16-s delay; cho, choice signal of white image. Bottom trace, events chart used in *b* and *d*. *b*, Raster recordings of impulse discharges and spike-density histogram obtained from a cell in trials where a 'learned' picture was used as a sample stimulus. These trials were originally separated by intervening trials of other 'learned' or 'new' sample stimuli, and were sorted and collected by off-line computation. Bin width, 200 ms. *c*, Frequency distribution of average firing rate measured during the delay period following 97 'learned' sample pictures in the same cell. The black column indicates the response shown in *b*. $\downarrow$ max and $\downarrow$ med, maximum and median value in the distribution. The selectivity of this neuron to these learned pictures, measured by the kurtosis of the response distribution, was 11.1. The kurtosis (ref. 13) is defined as: $N^{-1}(\sum_{i=1}^N ((x_i - \bar{x})/s)^4) - 3$, where x_i is the response to the i th picture, $\bar{x}$ and s are the mean and standard deviation of the distribution, N is the total number of pictures tested. *d, e*, Similar to *b, c* but 'new' pictures are used as sample stimuli in the same cell. The black column in *e* represents the response illustrated in Fig. 1*d*. The kurtosis of the response distribution shown in *e* was 5.2.

Methods. In our modified delayed matching-to-sample (DMS) task¹⁰, sample and match stimuli were successively presented on a video monitor, each for 0.2 s at a 16-s delay interval. The trials with identical or non-identical match stimulus were mixed randomly. Drops of fruit juice were administered as a reward. After the monkey had learned the DMS rule to an 85% performance level, he was over trained for two weeks with a fixed set of 97 fractal patterns ('learned stimuli') presented in a fixed sequence according to an arbitrarily attached number (serial position number, SPN). The same training was carried out throughout the search period. During the unit recording, a new set of patterns ('new stimuli') was created for each neuron and also formally given SPNs. The sample and match stimuli were then selected at random from the 'learned' and 'new' patterns, independent of the SPNs. The monkeys' performance level did not differ significantly for 'learned' and 'new' stimuli (85–100% correct). Error trials were excluded from the analysis in this report.



produce deficits in visual recognition of objects^{3–9}. We previously reported that in the anterior ventral temporal cortex of monkeys, individual neurons have a sustained activity that is highly selective for a few of the 100 coloured fractal patterns used in a visual working-memory task¹⁰. Here I report the development of this selectivity through repeated trials involving the working memory. The few patterns for which a neuron was conjointly selective were frequently related to each other through stimulus–stimulus association imposed during training. The results indicate that the selectivity acquired by these cells represents a neuronal correlate of the associative long-term memory of pictures.

Two adult monkeys (*Macaca fuscata*) were trained to perform a visual memory task¹⁰ (Fig. 1*a*). Each monkey memorized a sample stimulus during a delay period and then had to decide whether a second stimulus was the same as the sample. A set of 97 colour patterns was generated by a fractal algorithm with a 32-bit seed of random numbers¹¹; the set was repeatedly used during an over-training session ('learned stimuli'). While extracellular neural discharges were recorded using standard physiological techniques¹², a sample stimulus was selected not only from the 97 learned patterns but also from a new set of 97 patterns ('new stimuli'). Different sets of new stimuli were created for each neuron using the same algorithm but a different seed. The 'learned' and 'new' stimuli were used at random.

A few of the 97 learned stimuli reproducibly activated a particular neuron with high-frequency sustained discharges during the delay period of the task¹⁰ (Fig. 1*b* and *c*). The optimal set of stimuli varied from cell to cell. By contrast, the 97 new patterns produced only weak delay responses (Fig. 1*d* and *e*). No systematic tendency was found for the delay discharges to a new sample stimulus to become augmented or suppressed with

up to 10 repetitions (Fig. 1*d*). The distribution of the delay discharge rate to the new stimuli (Fig. 1*e*) lacked the small population of high-frequency responses (>2 impulse s^{-1}) that characterized the distribution of responses to the learned stimuli (Fig. 1*c*).

Of the 206 neurons recorded in the anterior ventral temporal cortex of the over-trained monkeys, 57 neurons exhibited shape-selective delay discharges as described previously¹⁰, and 17 of the 57 cells were successfully tested with both learned and new stimuli. In these 17 cells, the maximum delay discharge for learned stimuli (for example the arrow in Fig. 1*c*) was always larger than that for new stimuli (for example the arrow in Fig. 1*e*), and the difference was highly significant ($P < 0.001$, Wilcoxon test). When the selectivity of a neuron to a set of 97 patterns was represented by the sharpness of the response distribution (the kurtosis of the distribution¹³, see Fig. 1*c* legend), the selectivity to the learned patterns was almost always (15/17) higher than that to the new patterns ($P < 0.005$). Despite the presence of a few effective pictures in the learned stimuli, there was no difference between the median responses to the learned and new stimuli (for example the double arrows in Fig. 1*c* and *e*) for the 17 cells ($P > 0.5$).

The effectiveness of the learned stimuli cannot be due to any special feature of their geometric patterns for two reasons: (1) the set of learned stimuli and each set of new stimuli were artificially generated by the same fractal algorithm¹¹, thus they belonged to the same class of geometric patterns; (2) different sets of learned stimuli assigned to each monkey yielded identical results. Therefore the sharpness of the response selectivity of these neurons to the learned patterns was most likely to have been formed throughout training.

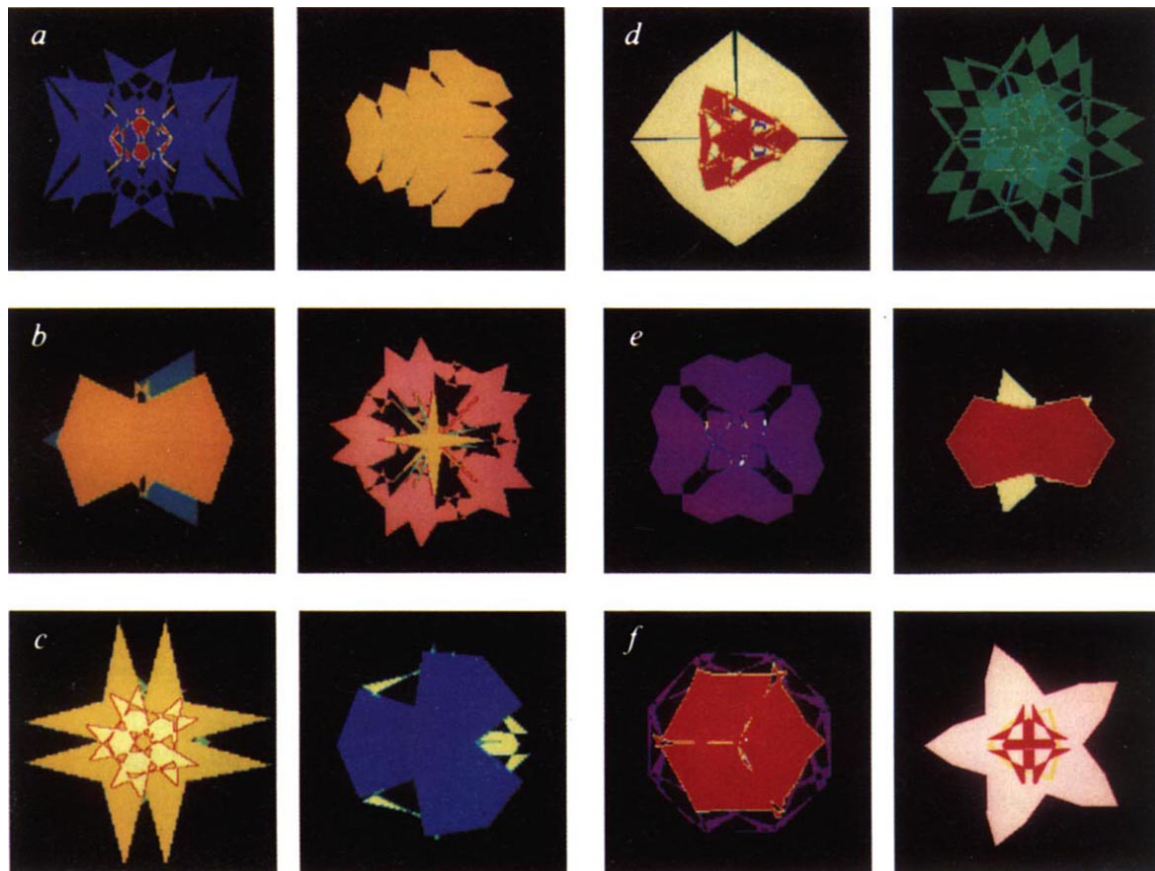
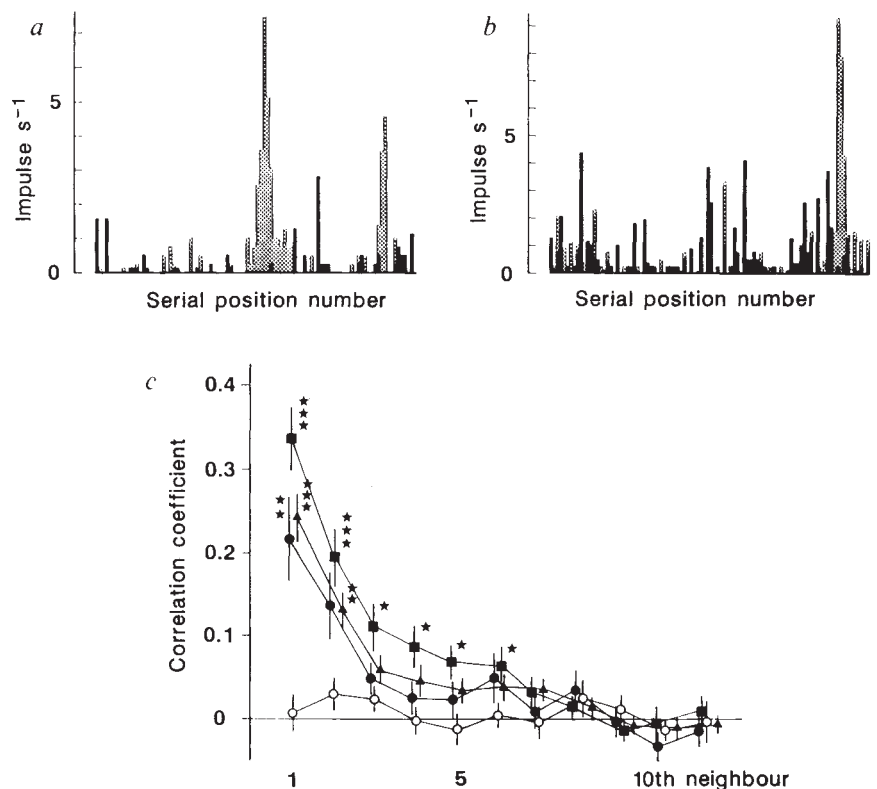


Fig. 2 Coloured fractal patterns that produced the two strongest delay discharges in learned stimuli for six different cells (*a-f*). The two optimal stimuli shown for each neuron have no similarities in their geometric patterns.

Fig. 3 Stimulus-stimulus association among the learned fractal patterns. *a*, Average delay discharge rate (impulse s^{-1}) for each sample stimulus in a cell against serial position number (SPN) of the stimuli (see Fig. 1 legend). Stippled columns, learned stimuli; black columns, new stimuli. *b*, As *a*, but for a different cell. *c*, Autocorrelograms of the delay discharge rate along the SPN of the stimuli. ● and ○, Average autocorrelogram for the learned and new stimuli in the 17 cells; ▲, that for the learned stimuli in the 57 cells; ■, that for the learned stimuli in the 28 cells for which the nearest-neighbour correlation along the SPN was significant ($P < 0.05$) according to Kendall's test. Error bars, standard errors. Three vertical asterisks $P < 0.001$, two vertical asterisks, $P < 0.01$, one asterisk, $P < 0.05$, according to the Kolmogorov-Smirnov test in comparison with the value for new stimuli.



Is it possible that training determines not only how sharply the effective learned patterns are represented in each neuron (as shown above), but also which patterns are conjointly chosen as the few optimal stimuli. I examined the geometric similarities between the optimal stimuli of a cell, and found that the stimuli were often completely different (Fig. 2), indicating that there is a non-geometric criterion for choice of these patterns as the optimal stimuli of a cell. I then examined the effect of a fixed-order presentation of the patterns during the training session according to an arbitrarily attached serial position number (SPN; see Fig. 1 legend). I expected that if the consecutively presented patterns tended to be associated together and if the association was fixed in the choice of effective patterns for a cell, the effective patterns would be correlated along the SPN, in spite of a random presentation of the stimuli during the unit-recording session.

Figure 3 shows that the effective responses to the learned stimuli do indeed cluster along the SPNs (*a* and *b*, stippled columns). The clustering was not due to an artefact in the testing procedure because the responses simultaneously obtained from the new stimuli were not clustered (black columns). In the 17 cells that were tested with both learned and new stimuli, the responses to the learned stimuli were significantly correlated (Fig. 3c, ●) in the nearest-neighbour of the SPNs, compared with the responses to the new stimuli (○) ($P < 0.01$, Kolmogorov-Smirnov test). The 57 cells tested by the learned stimuli had similar correlated responses along the SPN (▲). The nearest-neighbour correlations for the learned stimuli differed from cell to cell, and were significant ($P < 0.05$) in 28 of the 57 cells according to Kendall's correlation test¹³, whereas those for the new stimuli were not significant in any of the 17 cells. The autocorrelogram along the SPNs in these 28 cells (■) differed significantly up to the sixth neighbour from that of the new stimuli, but not from the seventh neighbour onwards ($P > 0.2$). This implies that the cells could associate pictorial patterns for the six consecutive neighbours during learning.

In the primate inferior temporal (IT) cortex, neural discharges selective to specific complex objects have been reported for hands^{6,14,15}, faces^{15,16}, Fourier descriptors¹⁷ and geometric patterns used in a discrimination task¹⁸. However, the selectivity of these sensory responses seemed to be organized along the patterns' geometric similarity¹⁴⁻¹⁸. The neurons reported here were selective for the artificial fractal patterns and had two distinct properties compared with other shape-selective IT neurons. First, these neurons were active in the delay period of the working-memory task, that is, when the visual stimulus was off and the monkey memorized it. Second, most other sensory neurons were located posterior to the area explored here, TE_{av} (ref. 19) (or TE₁-TE₂ (ref. 20)). TE_{av} has been anatomically designated as the last link from the visual system to the limbic memory systems^{7,19,21,22}, especially the hippocampus²³. It is tempting to assume that this interaction with the limbic system is important in organizing neuronal selectivity for the association of geometrically dissimilar pictures.

The work presented here shows directly that the responses of a single IT neuron reflect an experimentally controlled association between a temporally related set of stimuli. There are indirect clues indicating that there is such association in daily life; in the polysensory areas deep within the superior temporal sulcus (mainly area TPO)²⁰, which receive a projection from the IT cortex²⁴, some cells were selective for a person's different perspective views (face or body movement)²⁵, which might be interpreted as an associative experience between the perspective views that are visually distinct but often temporally related. Some face neurons did indeed change their responses when new faces were repeatedly shown²⁶. Visual associative agnosia^{3,27,28} in humans may result from selective loss of the neurons, thereby encoding the association between visually dissimilar objects.

The author thanks Professor Masao Ito for his encouragement.

Received 5 July; accepted 19 September 1988.

- Anderson, J. R. & Bower, G. H. *Human Associative Memory* (Lawrence Erlbaum Associates Hillsdale, 1980).
- Penfield, W. & Perot, P. *Brain* **86**, 595-697 (1963).
- McCarthy, R. A. & Warrington, E. K. *J. Neurol. Neurosurg. Psychiatr.* **49**, 1233-1240 (1986).
- Kimura, D. *Arch. Neurol.* **8**, 48-55 (1963).
- Milner, B. *Neuropsychologia* **6**, 191-209 (1968).
- Gross, C. G. in *Handbook of Sensory Physiology* Vol. 8, part 3B (ed. Jung, R.) 451-482 (Springer, Berlin, 1972).
- Mishkin, M. *Phil. Trans. R. Soc. B298*, 85-95 (1982).
- Gaffan, D. & Weiskrantz, L. *Brain Res.* **196**, 373-386 (1980).
- Sahgal, A., Hutchison, R., Hughes, R. P. & Iverson, S. D. *Behav. Brain Res.* **8**, 361-373 (1983).
- Miyashita, Y. & Chang, H. S. *Nature* **331**, 68-70 (1988).
- Miyashita, Y. in *Competition and Cooperation of Neural Nets* Vol. 2 (ed. Arbib, M.) (Springer, Berlin, in the press).
- Miyashita, Y. & Nagao, S. *J. Physiol.* **351**, 251-262 (1984).
- Snedecor, G. W. & Corcoran, W. G. *Statistical Method* (Iowa University Press, Iowa, 1980).
- Gross, C. G., Rocha-Miranda, C. E. & Bender, D. B. *J. Neurophysiol.* **35**, 96-111 (1972).
- Desimone, R., Albright, T. D., Gross, C. G. & Bruce, C. J. *Neurosci.* **4**, 2051-2062 (1984).
- Rolls, E. T. & Baylis, G. C. *Brain Res.* **65**, 38-48 (1986).
- Schwartz, E. L., Desimone, R., Albright, T. D. & Gross, C. G. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5776-5778 (1983).
- Sato, T., Kawamura, T. & Iwai, E. *Expl. Brain Res.* **38**, 313-319 (1980).
- Turner, B. H., Mishkin, M. & Knapp, M. J. *comp. Neurol.* **191**, 515-543 (1980).
- Seltzer, B. & Pandya, D. N. *Brain Res.* **149**, 1-24 (1978).
- Herzog, A. G. & Van Hoesen, G. W. *Brain Res.* **115**, 57-69 (1976).
- Van Hoesen, G. W. & Pandya, D. N. *Brain Res.* **95**, 1-24 (1975).
- Yukie, M. *Neurosci. Res. Suppl.* **7**, S16 (1988).
- Jones, E. G. & Powell, T. P. S. *Brain* **93**, 793-820 (1970).
- Perret, D. I., Mistlin, A. J. & Chitty, A. J. *Trends Neurosci.* **10**, 358-364 (1987).
- Baylis, G. C., Hasselmo, M. E. & Rolls, E. T. *Soc. Neurosci. Abstr.* **13**, 1317 (1987).
- Hecaen, H., Goldblum, M. C., Masure, M. C. & Ramier, A. M. *Neuropsychologia* **12**, 447-464 (1974).
- Marshall, J. C. *Nature* **334**, 378 (1988).

Persistent protein kinase activity underlying long-term potentiation

Roberto Malinow*, Daniel V. Madison* & Richard W. Tsien*

Department of Cellular and Molecular Physiology, Yale University School of Medicine, New Haven, Connecticut 06510, USA

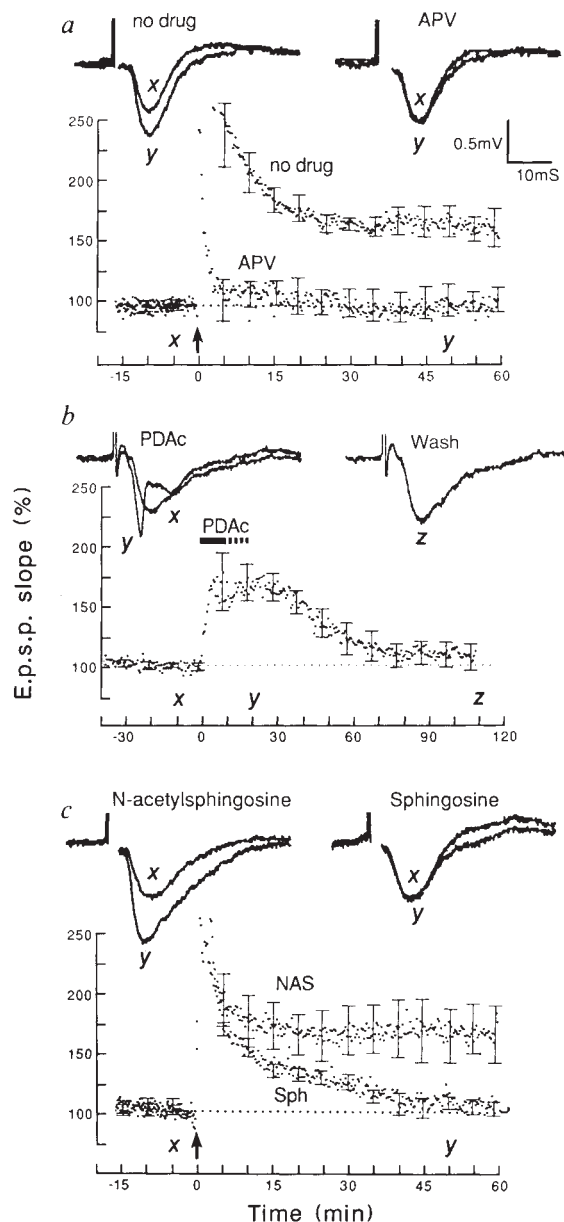
Long-term potentiation (LTP) of synaptic transmission in the hippocampus is a much-studied example of synaptic plasticity^{1,2}. Although the role of N-methyl-D-aspartate (NMDA) receptors in the induction of LTP is well established³⁻⁵, the nature of the persistent signal underlying this synaptic enhancement is unclear. Involvement of protein phosphorylation in LTP has been widely proposed⁶⁻¹⁵, with protein kinase C (PKC)^{6-8,10-12,14} and calcium-calmodulin kinase type II (CaMKII)^{9,13} as leading candidates. Here we test whether the persistent signal in LTP is an enduring phosphoester bond, a long-lived kinase activator, or a constitutively active protein kinase by using H-7, which inhibits activated protein kinases¹⁶ and sphingosine, which competes with activators of PKC (ref. 17) and CaMKII (ref. 18). H-7 suppressed established LTP, indicating that the synaptic potentiation is sustained by persistent protein kinase activity rather than a stably phosphorylated substrate. In contrast, sphingosine did not inhibit established LTP, although it was effective when applied before tetanic stimulation. This suggests that persistent kinase activity is not maintained by a long-lived activator, but is effectively constitutive. Surprisingly, the H-7 block of LTP was reversible; evidently, the kinase directly underlying LTP remains activated even though its catalytic activity is interrupted indicating that such kinase activity does not sustain itself simply through continual autophosphorylation (see refs 9, 13, 15).

We studied LTP in the Schaffer collateral/commissural-CA1 pathway in rat hippocampal slices. Following a conditioning tetanus, post-tetanic potentiation decayed within ~5 min and was followed by a more slowly decaying potentiation that disappeared within about 30 min (Fig. 1a; refs 4, 19). These com-

* Present address: Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, California 94305, USA.

Fig. 1 Effects of APV, sphingosine and N-acetylsphingosine on LTP and of phorbol-12,13-diacetate (PDAC) on synaptic transmission. Insets: representative records of extracellular excitatory postsynaptic field potentials (e.p.s.ps), obtained at a time marked by lower-case letter on time axis of graph. Graphs: maximal negative slope of rising phase of e.p.s.ps plotted against time²⁶. Averages of data from several slices are shown. The transient increase in e.p.s.p. slope immediately following the tetanus (post-tetanic potentiation) generally reached values >250% of control and are not displayed. *a*, Following tetanic stimulation (arrow), significant LTP occurred in control slices ($P < 0.01$, $n = 6$; see methods), but not in slices bathed in 50 μ M APV ($n = 5$). APV also blocked a slowly decaying component of potentiation. *b*, Application of 10 μ M PDAC produced significant synaptic enhancement ($P < 0.01$) which decayed back to a level not significantly elevated over baseline ($n = 6$). Duration of PDAC application varied from 10–20 min as indicated. *c*, Slices treated with sphingosine (Sph; 10 μ M) for 1–8 h before recording displayed no significant LTP but still showed slowly decaying potentiation and post-tetanic potentiation ($n = 8$). In slices identically treated with 10 μ M N-acetylsphingosine (NAS) for 1–8 h, LTP was significant ($P < 0.01$, $n = 8$).

Methods. Hippocampal slices (0.5 mm) were prepared by standard methods²⁷, submerged and superfused continuously with a modified Earle's solution containing (mM): NaCl, 119; KCl 2.5; MgCl₂, 1.3; CaCl₂, 2.5; NaH₂PO₄, 1.0; NaHCO₃, 26.2; glucose, 11 and gassed with 95% O₂/5% CO₂ (pH 7.4) at room temperature (22 °C). Sphingosine and N-acetylsphingosine were prepared in 100% ethanol, diluted 1000-fold in Earle's solution, and dispersed in a bath-type sonicator. PDAC was dissolved in dimethyl sulphoxide, diluted 1,000-fold and sonicated. Other drugs were dissolved directly in Earle's solution. In pre-treatment experiments (Figs 1c, 3a, c), slices were incubated in drug-containing solutions for 1–8 h before transfer to recording chamber and exposure to drug-free superfusate. These pre-treated slices received a tetanic stimulation within 30 min of their transfer from the incubation solution. E.p.s.ps were elicited every 15 s with bipolar stainless steel stimulating electrodes and recorded with a 3 M NaCl-filled glass microelectrode placed in stratum radiatum. LTP was elicited by a pair of stimulus trains, 100 Hz for 1 s, delivered ~25 s apart. E.p.s.ps were filtered at 1 kHz, digitized at 20 kHz and stored. In collected results from several slices, e.p.s.p. slopes from individual slices were normalized by their mean pre-tetanus values, time-registered with respect to the tetanic stimulation. Values of mean $\pm$ standard error of the mean were computed for each time point. Occurrence of LTP was judged by comparison of collected values 55–60 min after tetanic stimulation with control values 5–10 min before tetanic stimulation. $P < 0.05$ (Student's *t*-test) was taken as significant.



ponents of potentiation gave way to long-term potentiation, seen as a persistent (≥ 60 min) and significant enhancement of synaptic transmission above the pre-tetanus baseline. Blockade of NMDA receptors during the conditioning tetanus with DL-amino-phosphonovalerate (APV) prevented LTP^{4,5} and also eliminated the slowly decaying potentiation (Fig. 1a), which is consistent with the latter being produced by the application of NMDA^{4,19}.

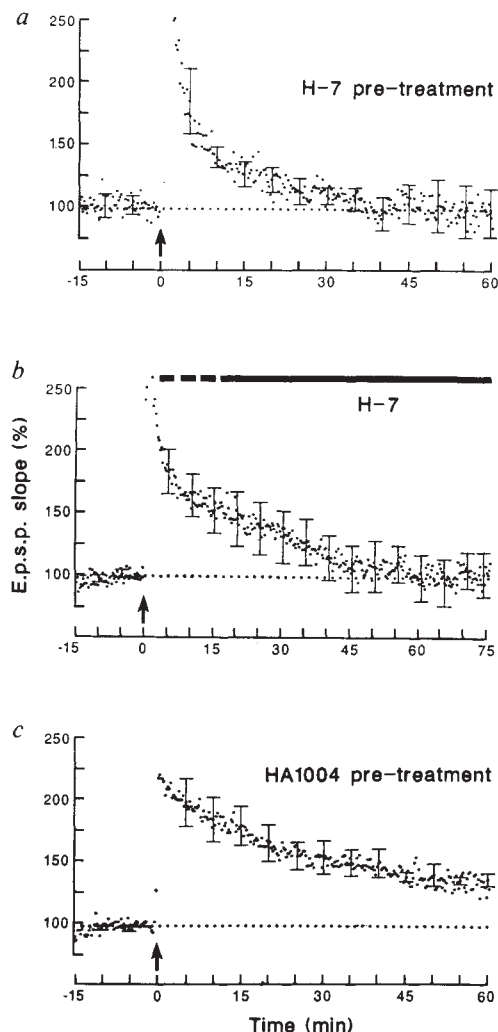
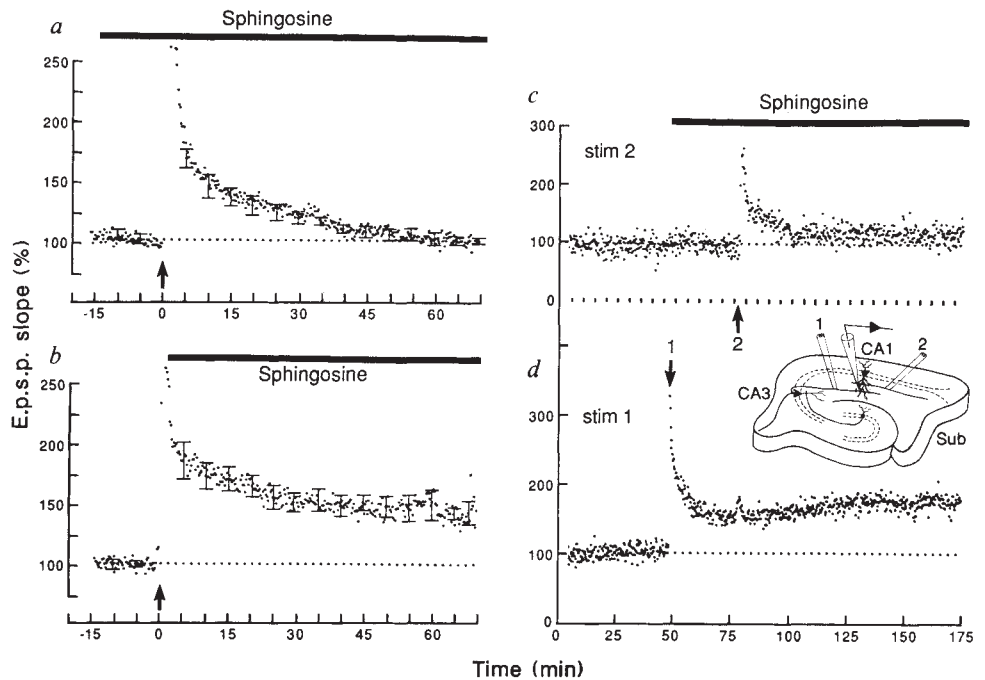
To investigate the role of protein kinases in LTP, we used compounds that are known to activate them or to prevent their activation. Phorbol esters, which activate PKC, produce a potentiation of synaptic transmission that resembles LTP (ref. 12). Here we show that the enhanced synaptic transmission decays back to baseline within one hour of the start of drug washout (6/6 trials). The reversal of synaptic potentiation, not seen in previous experiments¹², is probably due to improved perfusion of our recording chamber. Thus, although previous reports strongly suggest a role for PKC^{6–8,10–12,14}, we show that PKC activation by phorbol esters is not sufficient to establish LTP.

To test whether protein kinase activation is necessary for LTP, we studied the effects of sphingosine, which competes with diacylglycerol for PKC activation¹⁷ and also inhibits CaMKII

activation (ref. 18; M. B. Kennedy, personal communication) by competing with calmodulin¹⁸. Sphingosine did not reduce basal synaptic transmission (Fig. 2c, $n = 7$), but did block LTP, leaving post-tetanic potentiation and slowly decaying potentiation intact (Fig. 1c). However, when slices were pre-treated in the same way with N-acetylsphingosine, an analogue of sphingosine that is inactive against PKC (ref. 17) and only weakly active against CaMKII (ref. 18; M. B. Kennedy, personal communication), they showed normal LTP (Fig. 1c). Although sphingosine prevented LTP if applied as little as 15 min before tetanic stimulation, administration of sphingosine immediately following the tetanus failed to inhibit LTP, even though the exposure to the drug was much longer than 15 min (Fig. 2). The change in responsiveness to sphingosine was particularly convincing when postsynaptic responses to independent inputs were monitored in the same slice (Fig. 2c, d).

The effects of sphingosine were compared to those of H-7, a compound that inhibits the catalytic activity of activated protein kinases, including PKC (ref. 16) and CaMKII (M. B. Kennedy; A. B. Jefferson and H. Schulman, personal communications), by competing with ATP. To ensure maximal inhibition of kinase activity in this whole-tissue preparation, we used a relatively high concentration (300 μ M; at 50 μ M, H-7 gave similar but

Fig. 2 Sphingosine blocks LTP if applied before but not after tetanic stimulation. *a*, Collected data showing that tetanic stimulation produced no significant LTP when sphingosine was delivered by bath application beginning 15–30 min before the tetanus ($n=8$). *b*, Sphingosine applied immediately after the tetanus failed to block LTP as the e.p.s.p. remained significantly potentiated ($n=8$, $P<0.01$). *c*, *d*, Data from an individual slice. E.p.s.ps recorded at a common site in response to stimuli delivered at two locations (inset), stim 1 and stim 2. Stim 1 and stim 2 were each delivered every 15 s and staggered 7.5 s apart. Independence of stimulated presynaptic pathways was ensured by showing that tetanic stimulation at one site failed to potentiate transmission in response to the other stimulus, as illustrated. Sphingosine ($10\text{ }\mu\text{M}$) was added to the superfusate immediately after tetanic stimulation at stim 1 (downward arrow) and responses to this input (*d*) exhibited LTP. 30 min later, tetanic stimulation was applied at stim 2 (upward arrow) and responses in this pathway (*c*) failed to show LTP. This experiment is representative of results from six slices; the results were the same regardless of which electrode was tetanized first.

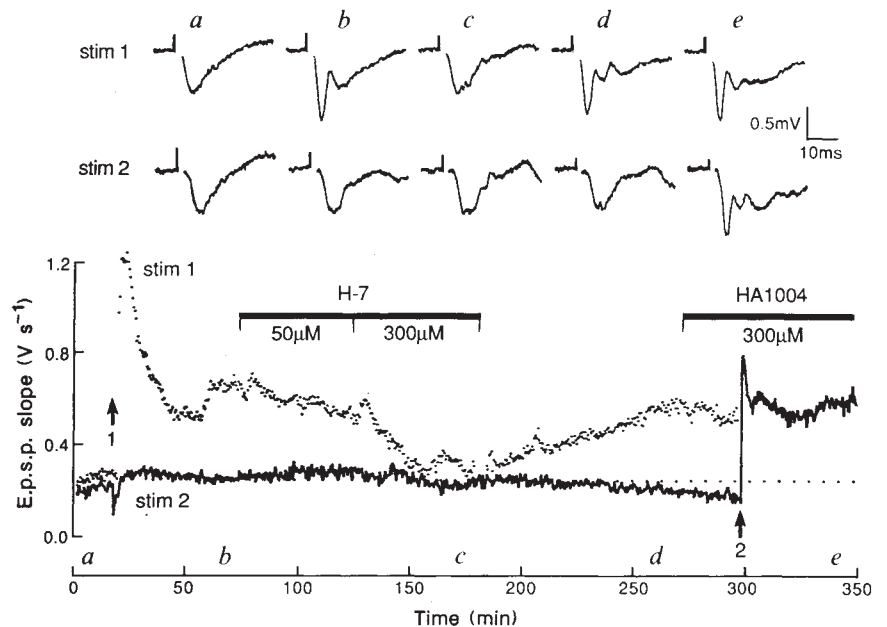


slower and more variable effects). Like sphingosine (Fig. 1*b*), H-7 prevented LTP when administered before the tetanus and left post-tetanic potentiation and slowly decaying potentiation intact (Fig. 3*a*). Unlike sphingosine, H-7 applied after the tetanus caused synaptic transmission to decay to the pre-tetanus basal level (Figs 3*b*, 4; see also ref. 10). There was complete inhibition of established LTP by H-7 ($92\pm4\%$, $n=38$) even when the drug was administered as late as three hours after the tetanus (the latest application tested). In contrast, H-7 had no effect on basal synaptic transmission in 22/31 experiments and little or no effect overall ($3.3\pm3.6\%$ depression, $n=31$). The selectivity of H-7 for LTP was most clear when monitoring independent inputs within a slice where LTP was reversed without change in basal transmission (11/14 experiments; see Fig. 4).

Does post-tetanic application of H-7 eliminate a persistent signal underlying the maintenance of LTP or merely block expression of the signal? As Fig. 4 illustrates, washout of H-7 can be followed by almost complete recovery of the original potentiation. Complete or partial recovery was seen in 11/13 slices, indicating that a persistent signal remains in the presence of H-7 and that only the expression of this signal is blocked. H-7 blocks several protein kinases with approximately equal potency¹⁶. To learn more about the target of H-7 action, we compared its effects with those of HA1004, a similar compound that is much less potent than H-7 with regard to PKC (ref. 16) but roughly equipotent against several other protein kinases¹⁶.

Fig. 3 LTP is blocked by H-7 applied either before or after tetanic stimulation, but not by HA1004. *a*, Collected responses from slices incubated with $300\text{ }\mu\text{M}$ H-7 for 1–8 h before recording (pre-treatment, $n=7$). *b*, Collected data ($n=7$) from slices superfused with $300\text{ }\mu\text{M}$ H-7 beginning 0 to 15 min after tetanic stimulation (dashed bar) and continuing to the end of the experiment (solid bar). No significant LTP was produced regardless of whether H-7 was applied before or after the tetanic stimulation. *c*, Collected data from slices incubated with $300\text{ }\mu\text{M}$ HA1004 for 1–8 h before recording (pre-treatment). In these slices, significant levels of LTP were produced by tetanic stimulation ($n=6$, $P<0.01$).

Fig. 4 H-7 applied post-tetanus reversibly blocks LTP, but has no effect on basal synaptic transmission, whereas HA1004 is not effective. E.p.s.p. slopes from a single recording site in response to stimuli alternating between two separate inputs (see inset, Fig. 2*d*). Following a tetanic stimulation at stim 1 (arrow 1), the slice was exposed to H-7 which had little effect at 50 μ M and inhibited LTP at 300 μ M but had no effect on the untetanzed pathway (stim 2). When H-7 was washed from the slice, synaptic transmission in the tetanized pathway recovered to its potentiated level. HA1004 did not reverse LTP in stim 1, nor did it prevent potentiation when stim 2 was subsequently tetanized (arrow 2). *a-e*, Representative recordings in response to stim 1 (upper row) and stim 2 (lower row), taken at times indicated along time axis below.



HA1004 did not prevent the induction of LTP (Fig. 3*c*) and was also much less effective than H-7 in inhibiting established LTP (Fig. 4).

We can draw several conclusions from these results: (1) protein kinase inhibitors can prevent LTP without significantly affecting basal transmission; (2) the pharmacological profile of LTP changes with respect to the inhibitors shortly after tetanic stimulation; (3) blockade of established LTP can be reversed; and (4) activation of PKC is not sufficient to induce LTP (although phorbol esters do enhance synaptic transmission¹²). These results provide new information about the persistent signal underlying LTP and its induction, maintenance and expression.

Induction and maintenance¹ of the persistent signal are distinguished by sphingosine, which blocks LTP if applied before but not after tetanic stimulation. Although similar in this way to APV (refs 3, 4), sphingosine does not prevent NMDA receptor function¹⁹, so there must be a sphingosine-sensitive process in addition to NMDA receptor activation that is critical for inducing LTP but not for maintaining it. We have distinguished maintenance and expression for the first time by showing that H-7 inhibition of established LTP can be reversed by removal of the drug. This implies that a persistent signal underlying LTP can be maintained even during an interruption of its expression as potentiated transmission.

These conclusions about the nature of LTP persistence can be reached without detailed knowledge about the drugs' targets. Further inferences can be made about molecular mechanisms if the effects of sphingosine and H-7 are attributed to their respective inhibitory actions on the activation and catalytic activity of protein kinases such as PKC and CaMKII, as supported by the relative ineffectiveness of N-acetyl sphingosine and HA1004. In this light, the observation that LTP remains H-7-sensitive for hours after the tetanus points to persistent kinase activity rather than a long-lasting phosphorylated substrate as the underlying mechanism of the synaptic enhancement. The loss of responsiveness to sphingosine applied after the tetanus implies that persistent kinase activity is not sustained by activators of PKC or CaMKII, such as diacylglycerol or calmodulin, against which sphingosine competes.

An attractive mechanism for explaining persistent kinase activity in the absence of activators is autophosphorylation^{9,13,15}. LTP recovers following removal of H-7, as if the kinase underly-

ing synaptic enhancement remains activated during inhibition of its catalytic activity and resumes activity once the inhibitor is removed. Our observations argue against simple schemes in which a single kinase maintains its own activity by autophosphorylation while also directly controlling synaptic strength through substrate phosphorylation. However, the results do not exclude models involving a cascade of protein kinases with very different sensitivities to H-7, or kinase dephosphorylation that is significantly slower than substrate dephosphorylation.

The fact that enhanced transmission is H-7-sensitive, but sphingosine-insensitive leads us to suggest that LTP is maintained by a protein kinase that is, in effect, constitutively active. A candidate with the appropriate pharmacological profile is PKM, the isolated catalytic domain of PKC^{20,21}, which might be derived from the parent enzyme by Ca²⁺-dependent proteolysis^{21,22}. Alternatively, the persistent kinase activity might arise from a long-lasting translocation of PKC to the surface membrane^{6,23,24}, or by some alteration of CaMKII, if the modified forms of these enzymes prove to be sphingosine-insensitive.

In each case, a key question is how the persistent kinase activity can be triggered by a brief tetanic stimulus. Our experiments show that a sphingosine-sensitive process (possibly activation of PKC) is necessary, although activation of PKC by phorbol esters is not sufficient. Similarly, NMDA receptor activation is necessary^{3,4,5} but not sufficient^{4,19} for LTP. Thus, LTP may arise from a combination of factors, including NMDA receptor activation and protein kinase activation, and inhibition of any of these factors may prevent this form of synaptic plasticity.

Sphingosine and related long-chain bases have been proposed as endogenous regulators of PKC and levels of these compounds are elevated in sphingolipidoses²⁵. Our results raise the possibility that these compounds normally modulate activity-dependent changes in synaptic transmission. Excessive inhibition of synaptic plasticity with pathologically elevated levels of these compounds might contribute to the mental retardation associated with lipid storage diseases.

We thank J. Potts for technical support, R. M. Bell for the generous gift of N-acetyl sphingosine, C. F. Stevens for the loan of equipment, L. K. Kaczmarek, M. B. Kennedy, C. R. Loomis, H. Rasmussen and H. Schulman for helpful discussions and R. L. Rosenberg for careful reading of the manuscript. This work was supported by PHS grants to R.W.T. and R.M. D.V.M.

is a Lucille P. Markey Scholar and this work was supported in part by a grant from the Lucille P. Markey Charitable Trust.

Received 19 July; accepted 19 September 1988.

- Bliss, T. V. P. & Lynch, M. in *Long-term potentiation: Mechanisms and Key Issues* (eds P. W. Landfield & S. A. Deadwyler) 1-69 (Alan R. Liss, New York, 1987).
- Nicoll, R. A., Kauer, J. A. & Malenka, R. C. *Neuron* **1**, 97-103 (1988).
- Collingridge, G. L. & Bliss, T. V. P. *Trends Neurosci.* **10**, 288-293 (1987).
- Collingridge, G. L., Kehl, S. J. & McLennan, H. J. *Physiol.* **334**, 33-46 (1983).
- Harris, E. W., Ganong, A. H. & Cotman, C. W. *Brain Res.* **323**, 132-137 (1984).
- Akers, R. F., Lovinger, D., Colley, P., Linden, D. & Routtenberg, A. *Science* **231**, 587-49 (1986).
- Akers, R. F. & Routtenberg, A. *Brain Res.* **334**, 147-151 (1985).
- Hu, G.-Y. *et al. Nature* **328**, 426-429 (1987).
- Lisman, J. E. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3055-257 (1985).
- Lovinger, D. M., Wong, K. L., Murakami, K. & Routtenberg, A. *Brain Res.* **326**, 177-183 (1987).
- Lynch, M. A., Clements, M. P., Errington, M. L. & Bliss, T. V. P. *Neurosci. Lett.* **84**, 91-296 (1988).
- Malenka, R. C., Madison, D. V. & Nicoll, R. A. *Nature* **319**, 774-776 (1986).
- Miller, S. G. & Kennedy, M. B. *Cell* **44**, 861-870 (1986).
- Reyman, K., Frey, U., Jork, R. & Matthies, H. *Brain Res.* **440**, 305-314 (1988).
- Schwartz, J. H. & Greenberg, S. M. *A. Rev. Neurosci.* **10**, 459-476 (1987).
- Hidaka, H., Inagaki, M., Kawamoto, S. & Sasaki, Y. *Biochemistry* **23**, 5036-5041 (1984).
- Hannun, Y. A., Loomis, C. R., Merrill, A. H. & Bell, R. M. *J. biol. Chem.* **261**, 12604-12609 (1986).
- Jefferson, A. B. & Schulman, H. *J. biol. Chem.* (in the press).
- Kauer, J. A., Malenka, R. C. & Nicoll, R. A. *Nature* **334**, 250-252 (1988).
- Inoue, M., Kishimoto, A., Takai, Y. U. & Nishizuka, Y. *J. biol. Chem.* **252**, 7610-7616 (1977).
- Pontremoli, S. *et al. J. biol. Chem.* **261**, 8309-8313 (1986).
- Staubli, U., Larson, J., Thibault, O., Baudry, M. & Lynch, G. *Brain Res.* **444**, 153-158 (1988).
- Akers, R. F. & Routtenberg, A. *J. Neurosci.* **7**, 3976-3983 (1987).
- Alkon, D. L. & Rasmussen, H. *Science* **239**, 998-1005 (1988).
- Hannun, Y. A. & Bell, R. M. *Science* **235**, 670-674 (1987).
- Alger, B. E. & Teyler, T. J. *Brain Res.* **110**, 463-470 (1976).
- Alger, B. E. & Nicoll, R. A. *J. Physiol.* **328**, 105-123 (1982).

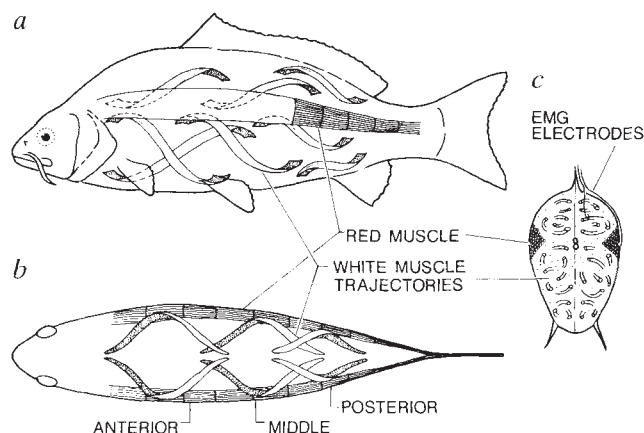


Fig. 1 Longitudinal view (a), dorsal view (b) and cross-section (c) of carp. The red muscle represents a thin sheet of muscle just under the skin which extends to a depth of only 10% of the distance to the backbone (the cross-section of the red muscle is exaggerated for illustrative purposes). Because the red fibres run parallel to the body axis (a, b) we were able to use simple beam theory to derive a straightforward relationship between sarcomere length on the convex and concave sides of the fish and the curvature of the spine:

$$\lambda_{\text{convex}} = \lambda_0(r+a)/r; \quad \lambda_{\text{concave}} = \lambda_0(r-a)/r,$$

where λ is sarcomere length, λ_0 the resting sarcomere length, a the half-thickness of the fish and r the radius of curvature at the spine. We verified this relationship by measuring the sarcomere length of muscle fibres from carp fixed in shapes found during swimming. The radius of curvature of the spine must be calculated from the average of the curvatures of the left and right sides of the fish because the spine is not visible in motion picture films. This involved digitizing the silhouettes of the left and right sides of the fish and determining each curvature with a seven-point moving-curve fit. The velocity of shortening (V) is given by the slope of the sarcomere length-time graph and was estimated with a five-point parabolic fit. The velocity of shortening of the white fibres was estimated by dividing V of the red fibres by four. The trajectories of the white muscle fibres shown in a and b are based on Alexander's¹³ description. The white fibres lie closer to the median plane than the red ones, and they run helically rather than parallel to the long axis of the body. Consequently, according to the calculations of Alexander¹³, they shorten by only $\sim 1/4$ as much as the red ones for a given curvature change of the body. Measurements of sarcomere lengths of white fibres from fixed fish are consistent with this calculation. Sarcomere length excursions were determined at three places along the fishes' body: an anterior position (38% of the distance from snout to tail), a middle position (52% of the distance from snout to tail) and a posterior position (68% of the distance from snout to tail). Placement of electromyography (EMG) electrodes used to determine the activity of the red and white muscle are shown in c. Note that the so-called pink muscle, a thin sheet interposed between the red and white, is not indicated here. It has been excluded from this analysis because its physiology and recruitment pattern has not been adequately characterized. Its exclusion, however, does not influence the analysis of the red and white muscle, which is based on their respective V/V_{max} values.

In vitro measurements of the mechanical and energetic properties of muscle are not adequate to demonstrate the advantage of having different fibre types; the velocity at which they shorten *in vivo* must be known. For instance, both mechanical power and efficiency are equal to zero at a $V/V_{\text{max}} = 0$ and $V/V_{\text{max}} = 1$, and reach a maximum at a $V/V_{\text{max}} = 0.2-0.3$ (refs 2-4). To test the advantage of having different fibre types it is necessary to measure the V/V_{max} of the slow and fast fibres both during slow speeds of locomotion and during maximal movements. Measurements of V/V_{max} require separate determinations of (1) the V at which the muscle fibres shorten during

Why animals have different muscle fibre types

Lawrence C. Rome*, Roel P. Funke†, R. McNeill Alexander‡, Gordon Lutz*, Hugh Aldridge‡, Frank Scott† & Marvin Freadman§

* Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania, 19104, USA and the Marine Biological Laboratories, Woods Hole, Massachusetts 02543, USA

† Department of Zoology, University of Tennessee, Knoxville, Tennessee 37996, USA

‡ Department of Pure and Applied Biology, University of Leeds, Leeds LS2 9JT, UK

§ Boston University Marine Program, The Marine Biological Laboratories, Woods Hole, Massachusetts 02543, USA

Animals have different muscle fibre types: slow fibres with a low maximum velocity of shortening (V_{max}) and fast fibres with a high V_{max} . An advantage conferred by the use of different fibre types during locomotion¹ has been proposed solely on the basis of their *in vitro* properties. Isolated muscle experiments show that force generation, mechanical power production and efficiency are all functions of V/V_{max} , where V is the velocity of muscle shortening. But it is not known whether animals actually use the different fibres at shortening velocities that are optimal for mechanical power production and efficiency. Here we compare the V of muscle fibres during locomotion with their V_{max} . This comparison shows that during slow locomotion, the slow fibres shorten at a velocity that gives peak mechanical power and efficiency and the fast fibres shorten at their optimal velocity when powering maximal movements. Our results also show that maximal movements are impossible without fast fibres because the slow ones cannot shorten rapidly enough.

Fig. 2 Sarcomere excursion, tail-beat frequency, and velocity of muscle shortening as a function of swim speed. Data are shown for one of the four fish examined in this study. At 20 cm s^{-1} the sarcomere length excursion was smaller in the anterior than in middle and posterior regions of the fish. As speed increased, the sarcomere excursion of the anterior fibres increased while that of the middle and posterior stayed the same. At the maximum speed the fish can swim before recruiting white muscle (40 cm s^{-1}), the sarcomere length excursion is the same at the three positions on the fish. Δ , Posterior; $\bullet$, middle; $\circ$, anterior.

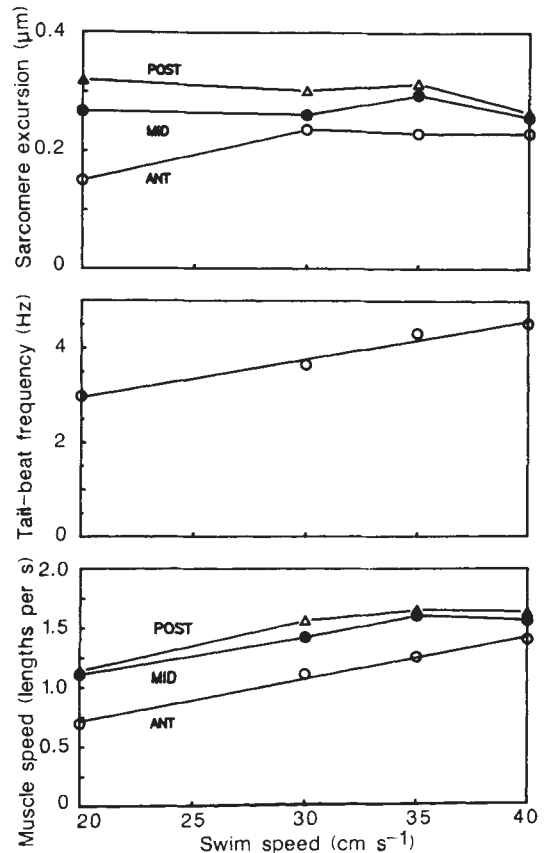
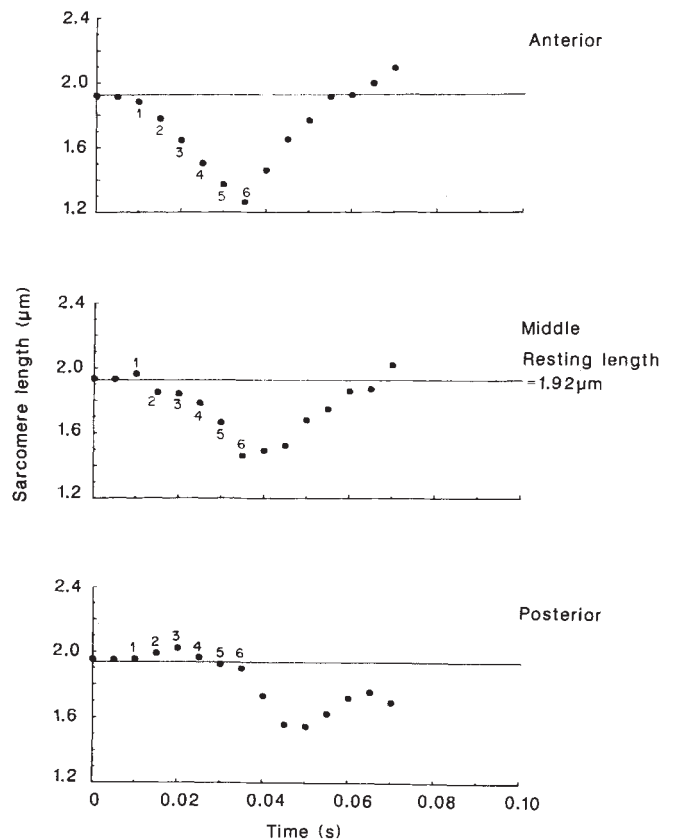
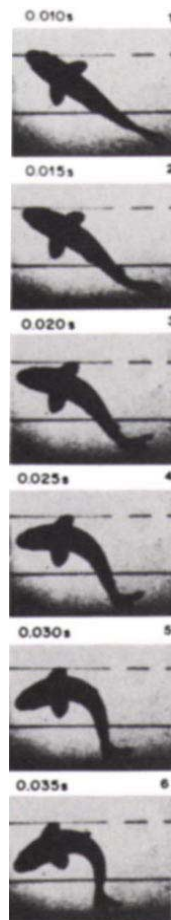


Fig. 3 The startle response of a carp. A resting carp received a 100 ms, 150 Hz sound pulse through an underwater speaker in the aquarium about 30 cm from the fish. The response was filmed at 200 frames per second and six consecutive frames (separated by 5 ms) are numbered and shown on the left. The numbers and times given on the photographs correspond to data points on the graphs of sarcomere length of red muscle as a function of time, shown on the right. Sarcomere lengths are shown for the three positions along the fish. It can be seen that the sarcomere shortening in the anterior region precedes that in the middle and posterior regions. Fibres appear to shorten to extremely short lengths of $\sim 1.27 \mu\text{m}$. On average, the velocity of shortening (the slope of the sarcomere length-time graph) was the same for all three positions, but the sarcomere excursion was greater in the anterior and middle positions than it was in the posterior. We stress that the sarcomere length changes and velocities presented for the startle response are apparent values (calculated from the curvature of the fishes' backbone). This movement is driven by the white muscle which causes the distance between the red fibre attachments (myosepta) to decrease rapidly. As the red muscle fibres cannot shorten fast enough by themselves to remain taut, there are two possible consequences. First, the shortening of the white fibres forces the thick and thin filaments of the red muscle to be pushed past each other at a much higher velocity than their V_{max} , thereby reducing the actual sarcomere length to the calculated value of $1.27 \mu\text{m}$. Second, the red muscle fibres may simply buckle and fold like a rope whose ends have been brought close together. At this time we do not know which possibility is correct.



locomotion; (2) the $V_{\max}$ of the different fibre types, and (3) which muscle fibre types are active at a given speed of locomotion. Of these three parameters, it is the last which most severely limits the selection of an animal model: carp (*Cyprinus carpio*) provide a good model because of the anatomical separation of their different muscle fibre types (Fig. 1). The activity of the different fibre types in carp can be monitored by electromyography^{5,6}. Electromyography and glycogen-loss experiments show that at low speeds (20–40 cm s⁻¹) the red muscle fibres alone are active, and at higher speeds the white muscle fibres are recruited as well.

The force-velocity relationship was measured on bundles of red fibres using the force clamp method⁷ over roughly the same sarcomere length excursions (2.05–1.82 μ m; see determination of V below) that the muscle underwent during swimming. The bundles (30–80 red muscle fibres) were maximally activated by electrical stimulation (200 pulses per s) in the presence of 1 mM caffeine and 10⁻⁵ g ml⁻¹ eserine. The mean $V_{\max}$ was 4.65 muscle lengths per second ($\pm$ s.e. = 0.55, $n = 5$) at 15 °C.

$V_{\max}$ was measured on single 'skinned'⁸ white muscle fibres at 15 °C using the slack test method^{7,9}. Maximal activation was achieved in a solution containing a free calcium concentration of 3.2 μ M and a free magnesium concentration of 1 mM, at an ionic strength of 225 mM. The mean $V_{\max}$ was 12.88 muscle lengths per second ($\pm$ s.e. = 0.5, $n = 6$).

The sarcomere length excursion and velocity of shortening of red and white muscle fibres were determined from high-speed motion pictures taken at 200 frames per second (see Fig. 1 legend). To determine V of the fibres during slow locomotion, films were taken of four carp swimming in a water treadmill at speeds (20–40 cm s⁻¹) and temperature (15 °C) where only the red muscle fibres were active. During swimming at 40 cm s⁻¹, sarcomeres in the red fibres underwent an excursion of about 0.27 μ m (s.e. = $\pm$ 0.01, $n = 9$) centred around the resting length of 1.92 μ m (that is, from 1.79 to 2.06 μ m). Figure 2 shows that the sarcomere length excursion was nearly independent of swim speed in the middle and posterior regions of the fish, but was lower in the anterior region of the fish at low swim speeds.

As the swim speed increased from 20 to 40 cm s⁻¹, we found that tail-beat frequency increased as well. Thus the fishes' muscles had to undergo the same sarcomere length change in a shorter period of time, resulting in an increase in the muscle-shortening velocity (Fig. 2). On average, there was no difference in shortening velocity (1.67 muscle lengths per second, s.e. = $\pm$ 0.13, $n = 4$) between the anterior, middle, and posterior positions of the fish at the highest speed the fish could swim before recruiting their white muscle. At the lowest swim speeds, however, the velocity of the anterior fibres was significantly smaller than that of the middle and posterior because of their smaller sarcomere length excursion (at 20 cm s⁻¹, anterior fibres shortened at 0.73 muscle lengths per second, s.e. = $\pm$ 0.04, $n = 4$ versus 1.08 for the middle and posterior fibres). Thus at slow swim speeds, the red muscle shortens at velocities from 0.73 to 1.67 muscle lengths per second. As can be seen from Fig. 4, this corresponds to a velocity range at which maximum power is generated by the muscle. At these same swim speeds, the white muscle, if it were active, would be shortening at only a quarter of the velocity of the red, or 0.18 to 0.42 muscle lengths per second (see Fig. 1 legend for the basis of the calculation).

To determine the effectiveness of the red and white muscle to power maximal movement, we photographed the 'startle response'. As can be seen in Fig. 3, the fish undergoes a sharp 'C' bend in a short period of time. Muscle shortening is far greater than that undergone during normal swimming, and is much more rapid. Red muscle fibres appear to shorten to a sarcomere length of 1.28 μ m (s.e. = $\pm$ 0.02, $n = 15$), whereas on the convex side they are stretched to about 2.56 μ m. This occurs in ~20 ms. On average, the velocity of shortening of muscle in the position of the red was 19.6 muscle lengths per second (s.e. = $\pm$ 0.84, $n = 15$), which far exceeds the $V_{\max}$ of the red

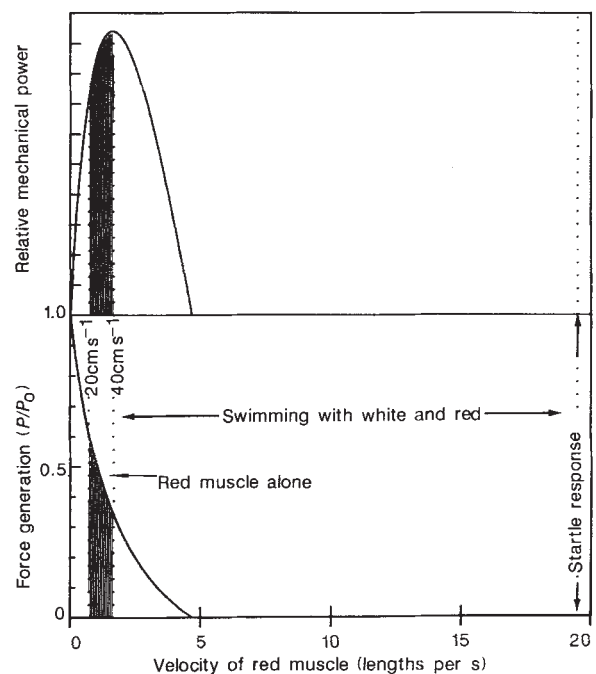


Fig. 4 Muscle function during locomotion. The average force-velocity (Hill equation constants: $V_{\max} = 4.65$ muscle lengths per second, $a/P_0 = 0.40$) and mechanical power curves obtained from the red muscle bundle experiments are graphed. The velocity of shortening at the position of the red muscle during slow swimming and maximal movements are shown on the graphs. The vertical dashed lines labelled 20 cm s⁻¹ and 40 cm s⁻¹ represent the V of the red muscle at these swim speeds. Between 20 and 40 cm s⁻¹ (shaded), the carp use their red muscle exclusively. Carp must recruit their white muscle when swim speed is faster than 40 cm s⁻¹ because the red muscle cannot generate enough mechanical power. The mechanical power needed for swimming increases as (speed)^{2.5}. It can be seen that at the maximum speed at which the fish can swim before recruiting their white muscle, the red muscle is shortening at a velocity where it is already generating its maximum mechanical power. As the fish swims faster, its red muscle would have to shorten at a higher velocity, and at a high shortening velocity the muscle would actually generate less mechanical power. To generate the additional mechanical power, the fish needs to recruit its white muscle, which results in rapid fatigue because of the low aerobic capacity of the white muscle. The dashed line shows the velocity at which the red muscle fibres must shorten to generate the startle response.

muscle (Fig. 4). The white muscle, because of its different orientation, needs to shorten at only 4.85 muscle lengths per second during a startle response.

Figure 4 shows how the red muscle is used during locomotion. The shaded portion represents the velocities at which fibres shorten while the fish is swimming between 20–40 cm s⁻¹ (speeds at which red muscle is used exclusively). The dashed line represents the shortening velocity in the position of the red muscle during the startle response. If the slow red muscle alone were to power this stage, it would have to generate the mechanical power for the movement while shortening at a velocity greater than its $V_{\max}$, which it clearly cannot.

Fast white muscle fibres, however, are active during the startle response^{10,11}. Their $V_{\max}$ of 12.88 muscle lengths per second would still not be fast enough to enable fibres in the position of the red muscle to generate the power for the escape response. Because of their different orientation, however, the white muscle needs to shorten at only 4.85 muscle lengths per second, which is about 0.38 $V_{\max}$. This is a position on its force-velocity curve where the white muscle would be expected to generate almost

its maximum mechanical power. It should be noted that if the red muscle were placed in the position occupied by the white, it would not be able to power the startle response because it would still have to generate mechanical power at a value of V (4.85 muscle lengths per second) which exceeds its $V_{\max}$ (4.65 muscle lengths per second). So fish have evolved fast white muscle fibres to power maximal movement, and it is both the higher $V_{\max}$ and the different orientation of the white muscle that permit maximal movement.

Although placing the white muscle in a helical orientation enables it to power a startle response, it makes it very ineffective at powering slow swim speeds. Recent information on dogfish muscle shows that maximum efficiency is achieved at a $V/V_{\max}$ of about 0.2–0.7 (ref. 12). This means that the range we found for $V/V_{\max}$ of red muscle (0.16 to 0.35) during slow swimming (20–40 cm s⁻¹) would represent maximal efficiencies. On the other hand, at these swim speeds, the white muscle would be shortening at 0.18 to 0.42 muscle lengths per second. For a muscle with a $V_{\max}$ of 12.88 muscle lengths per second, this corresponds to contractions that are nearly isometric ($V/V_{\max} = 0.01$ to 0.03). At this $V/V_{\max}$, the mechanical power generation and efficiency of the white muscle would be nearly zero. Thus, we can be reasonably certain that the red muscle can power slow swimming far more efficiently than white muscle.

In conclusion, our results provide evidence that slow muscle is used at low speeds of locomotion where it is most efficient, and show clearly that animals need fast muscle fibre types because the slow ones cannot shorten fast enough to generate maximal movement.

We thank Dr Steve Zottoli for help with the startle response experiments and Dr Andre Sosnicki for histochemistry and electron microscopy on the red fibre bundles. This research was supported by a Marine Biological Laboratories Summer Fellowship and Grants from the University of Tennessee, the Whitaker Foundation, and NIH to L.C.R.

Received 8 July; accepted 15 September 1988.

1. Goldspink, G. in *Mechanics and Energetics of Animal Locomotion* (eds Alexander, R. McN. & Goldspink, G.) 57–81 (Chapman and Hall, London, 1977).
2. Hill, A. V. *Scient. Prog. Lond.* **38**, 209–230 (1950).
3. Hill, A. V. *Proc. R. Soc. B* **159**, 319–324 (1964).
4. Kushermick, M. J. & Davies, R. E. *Proc. R. Soc. B* **174**, 315–353 (1969).
5. Rome, L. C., Loughna, P. T. & Goldspink, G. *Am. J. Physiol.* **249**, R272–R279 (1984).
6. Rome, L. C., Loughna, P. T. & Goldspink, G. *Science* **228**, 194–196 (1985).
7. Julian, F. J., Rome, L. C., Stephenson, D. G. & Striz, S. *J. Physiol.* **370**, 181–199 (1986).
8. Altringham, J. D. & Johnston, I. A. *J. Physiol.* **333**, 421–449 (1982).
9. Julian, F. J., Rome, L. C., Stephenson, D. G. & Striz, S. *J. Physiol.* **380**, 257–273 (1986).
10. Zottoli, S. J. *J. Exp. Biol.* **66**, 243–254 (1977).
11. Eaton, R. C., Lavender, W. A. & Wieland, C. M. *J. Comp. Physiol.* **144**, 521–531 (1981).
12. Curtin, N. A. & Woledge, R. C. *J. Physiol.* (in the press).
13. Alexander, R. McN. *J. Mar. Biol. Ass.* **49**, 263–289 (1969).

Preferential expression of the T-cell receptor $V_{\beta}3$ gene by Mls^c reactive T cells

Ryo Abe*, Melanie S. Vacchio*, Barbara Fox† & Richard J. Hodes*

* Experimental Immunology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA
† Rheumatology Division, University of Maryland Medical School, Baltimore, Maryland 21201, USA

The precursor frequency of T cells specific for any given foreign antigen is, in general, extremely low. Prominent exceptions to this rule are the T cells that are specific for foreign major histocompatibility complex (MHC) products or for products of the minor lymphocyte stimulatory (Mls) genes in the mouse which are present at high frequencies^{1,2}. Here, we report a striking overlap or cross-reactivity between the T cells specific for the protein antigen pigeon cytochrome *c* in association with E_α^kE_β^k and the set of T cells specific for Mls^c products. In addition, we demonstrate that the basis for this overlap is the predominant expression of one T-cell receptor (TCR) V_{β} gene, $V_{\beta}3$, by T cells that recognize Mls^c products. These results indicate the importance of specific TCR $\alpha\beta$ dimers in the recognition of Mls^c products and that positive or negative selection of T cells specific for Mls self-determinants may selectively alter the repertoire of T cells available for MHC-restricted recognition of foreign antigens.

To assess the relationship between the repertoire of T cells specific for conventional foreign antigens and T cells specific for Mls determinants, we examined antigen-specific T-cell clones for cross-reactive responses to Mls-disparate stimulator cells. None of the 20 clones specific for keyhole limpet haemocyanin, *Staphylococcus* nuclease, or MHC alloantigens were reactive to Mls^c stimulators (data not shown). In contrast, 10 of 15 cytochrome *c*-specific B10.A clones (refs 3–5) proliferated in response to Mls^c stimulators (Table 1; refs 6 and 7). To determine whether enrichment of cytochrome *c*-specific T cells would consistently select for Mls^c reactivity, B10.BR (Mls^b) T-cell lines specific for cytochrome *c* or purified protein derivative (PPD)

were generated. The cytochrome *c*-specific line, but not the PPD-specific line, responded strongly to C3H/HeJ (Mls^c) stimulators (Fig. 1).

We next determined whether the allogeneic determinants recognized by cytochrome *c*-reactive clones are in fact Mls^c products. The response patterns of all cytochrome *c*-reactive clones tested to BXH recombinant inbred mice [derived from the strains C57BL/6 (Mls^b) and C3H/HeJ (Mls^c)] and to progeny of a formal segregation analysis [(Mls^c × Mls^b) × Mls^b] were identical to primary Mls^c-specific responses by AKR/J (Mls^a) T cells (data not shown) and responses by Mls^c-specific clones⁸ (Table 2). These results demonstrate that the allogeneic determinants recognized by cytochrome *c*-specific clones are Mls^c gene products (or the products of closely linked genes).

The relationship between Mls^c reactivity and TCR use was first evaluated in a series of cytochrome *c*-specific clones. All nine clones expressing $V_{\beta}3$ were reactive to Mls^c, regardless of the V_{α} expressed; and only one of the six $V_{\beta}3$ non-expressing clones was reactive to Mls^c (Table 1). These results suggested that $V_{\beta}3$ (and to a lesser extent $V_{\alpha}11$) expression might be selectively associated with Mls^c reactivity. Consistent with this, $V_{\alpha}11$ - and $V_{\beta}3$ -specific probes hybridized with RNA from an anti-cytochrome *c* line that was highly cross-reactive to Mls^c, but did not hybridize detectably with messenger RNA from a PPD-specific line (Fig. 2). The association of TCR expression with Mls^c specificity was further examined by studying Mls^c-specific B10.BR T-cell clones^{7,8} which had been generated by repeated stimulation with Mls^c cells and which have no cross-reactivity to cytochrome *c* (data not shown). The $V_{\alpha}11$ probe did not hybridize with RNA from any of these clones, whereas messenger RNA from all three Mls^c-specific clones hybridized strongly with $V_{\beta}3$ (Fig. 2). As the $V_{\beta}3$ probe did not hybridize with RNA from Mls^a-specific clones, these results indicate that $V_{\beta}3$ expression correlates selectively with Mls^c reactivity and has a direct effect on T-cell responses to Mls^c.

Recently, preferential $V_{\beta}8.1$ (ref. 9) and $V_{\beta}6$ (ref. 10) use in Mls^a-reactive T cells has been reported, indicating that the $\alpha\beta$ TCR heterodimer may be important in Mls^a recognition. The results presented here demonstrate a similarly important role of the TCR in recognition of a distinct gene product, Mls^c (ref. 11), indicating that expression of $V_{\beta}3$ is sufficient to confer Mls^c reactivity on a T cell, independently of other segments in the expressed TCR β - and α -chains. The level of expression of

Table 1 Responses of cytochrome *c*-specific T-cell clones to Mls-disparate strains

Clones	TCR usage		Responses of clones to stimulators from						B10.A + Ag*
	V _α	V _β	AKR/J (k, a)‡	B10.BR (k, b)	C3H/HeJ (k, c)	CBA/J† (k, d)	A/J (a, c)	B10.A (a, b)	
B3	11	3	226	145	11,849	4,011	15,853	233	13,388
5C.C6	11	3	735	1,292	27,520	23,188	28,833	2,123	28,647
31P	11	3	314	309	11,386	29,723	9,354	526	7,749
1C.5	11	3	613	474	28,709	14,391	14,922	604	19,155
B10§	11	16	16,149	2,784	19,039	26,707	16,591	3,223	14,995
3.3	F4	1	494	1,127	328	1,821	1,498	2,095	10,532
D4	11	3	759	1,191	18,785	3,246	16,585	1,607	16,633
3D3	11	3	216	209	3,437	1,367	5,686	357	10,160
2C2	11	3	1,927	1,694	37,468	43,162	30,509	2,390	14,952
C10	11	15	496	797	977	261	832	1,811	11,134
F11	F4	1	627	477	825	190	989	1,011	23,323
F1.A2	11	1	375	912	1,075	399	1,143	784	18,028
3.C6	?*	3	15,007	24,480	69,662	62,307	70,883	26,829	39,211
D2	11	16	241	307	555	191	663	245	4,545
4.D11	?*	3	210	375	3,116	1,109	3,114	273	7,562

Mixed lymphocyte responses (MLRs) were set up as previously described⁸ with 1×10^4 cloned T cells as responders and 4×10^5 3,000-rad-irradiated splenic stimulator cells from mice which were injected with 200 μ l of goat anti-mouse immunoglobulin D antiserum to enhance Mls stimulation without altering the specificity of the response¹⁶. Each clone has been tested several times and representative results are shown. Uptake of ^3H by responder cells is expressed as the arithmetic mean of triplicate cultures; standard errors were generally less than 10% of the mean. Bold numbers are those which are significantly greater ($P < 0.01$ based on Student's *t*-test) than responses to syngeneic stimulators alone. The generation, characterization and TCR expression of the cytochrome *c*-specific clones used in these experiments have been described elsewhere (refs 4 and 5; L. Matis, personal communication). ND, not determined.

* 25 $\mu\text{g ml}^{-1}$ of pigeon cytochrome *c* (Sigma) was added to a mixture of 1×10^4 cloned T cells and 4×10^5 irradiated B10.A spleen cells. † CBA/J (Mls^d) expresses Mls^a and Mls^c determinants⁷. ‡ H-2, Mls. § Reactivity of clone B10 to AKR/J may represent cross-reactivity to Mls^a. ||, Clone 3.C6 is reactive to syngeneic MHC determinants in addition to an apparent incremental response to Mls^c-expressing stimulator cells. * Not V_α11.

Table 2 Mls^c specificity of cloned T-cell responses

Stimulator strain	Responding T cells					αMls ^c clone BC3C13*	αH-2 ^k response B10 T
	B3	Responses of α-cytochrome <i>c</i> clones 1C5	31P	2C2	5C.C6		
C3H/HeJ	43,921	37,431	35,363	54,421	41,688	40,543	109,301
B6	642	42,543 †	44,100 †	49,877 †	979	309	6,749
BXH 3	24,069	8,518	44,341	43,406	15,239	9,357	29,161
BXH 6	38,319	17,955	47,479	52,218	19,620	24,461	28,071
BXH 7	555	241	328	567	219	143	14,030
BXH 14	43,225	56,573	53,147	56,573	24,002	28,489	57,565
AKR/J	361	542	133	1418	146	125	15,577
B10.A	975	754	203	1078	496	102	35,121
B10.A + Ag§	31,413	36,625	37,054	46,899	32,166	283	ND
C3H/HeJ	12,299	12,683	13,721	10,955	11,337	10,681	26,289
AKR/J	1,135	677	968	1,351	1,846	571	100,158
B10.BR	861	449	1,017	1,163	1,668	459	15,860
(C3H × AKR)F ₁	13,239	16,289	20,048	35,933	33,523	37,634	126,963
(C3H × AKR) × B10.BR							
1	14,419	20,423	43,338	34,263	25,156	49,052	47,269
2	2,699	1,809 ‡	1,832 ‡	3,943 ‡	6,447	2,049	44,710
3	1,319	1,050	831	1,153	1,652	357	37,047
4	8,479	6,581	11,454	10,505	15,423	34,423	35,855
5	8,573	11,335	36,931	24,809	22,375	37,408	29,465
6	11,335	9,216	39,427	27,331	23,276	40,687	87,274
7	1,596	842	1,198	1,695	2,469	383	95,415
8	1,225	882	1,810	1,716	2,619	438	57,375
9	17,293	13,807	20,945	23,773	22,666	30,632	46,679
B10.A	1,032	984	1,136	1,862	1,591	ND	48,336
B10.A + Ag§	20,651	19,962	59,183	36,350	42,699	ND	ND

MLRs were set up with 3×10^5 unprimed nylon nonadherent spleen T cells or 1×10^4 cloned T cells as responders and 4×10^5 3,000-rad-irradiated splenic stimulator cells from mice which were injected with goat anti-mouse immunoglobulin D antiserum. Uptake of ^3H by responder cells are expressed as the arithmetic means of triplicate cultures, standard errors were generally less than 10% of the mean. Bold numbers are those which are significantly greater ($P < 0.01$ based on the Student's *t*-test) than responses to syngeneic stimulators alone. ND, not determined.

* B10.BR-derived Mls^c-specific clone⁸. † The cross-reactivity of these clones to I-A^b has previously been described⁵. § 25 $\mu\text{g ml}^{-1}$ of pigeon cytochrome *c* was added to a mixture of 1×10^4 cloned T cells and 4×10^5 irradiated B10.A spleen cells. ‡ $P < 0.05$.

Fig. 1 Responses of cytochrome *c* or PPD-specific T-cell lines to Mls-disparate stimulators. In (a) and (b), responder cells were 3×10^5 draining lymph node cells from B10.BR mice which had been immunized with pigeon cytochrome *c* peptide 81-104 plus complete Freund's adjuvant (CFA) (a) or with CFA alone (b). In (c) and (d), responders were 2×10^4 cells from lines which had been stimulated five times *in vitro* with pigeon cytochrome *c* (c, derived from the lymph node population used in a) or with PPD (d, derived from the lymph node population used in b). Responder cells were co-cultured with designated numbers of stimulator cells from anti-mouse immunoglobulin D injected mice: *—* AKR/J (H-2^k, Mls^a); ○—○, B10.BR (H-2^k, Mls^b); □—□, C3H/HeJ (H-2^k, Mls^c); or △—△, B10.D2 (H-2^d, Mls^b). ■ 1 μ M pigeon cytochrome *c* (a, c) or 25 μ g ml⁻¹ of PPD (b, d) was used with 4×10^5 irradiated B10.BR spleen cells for antigen-specific T-cell responses. Methods of MLRs were described in Table 1. Clones used were shown to be independent (non-identical) by analysis of their independent derivation, fine specificity and/or TCR α - and β -chain³⁻⁵.

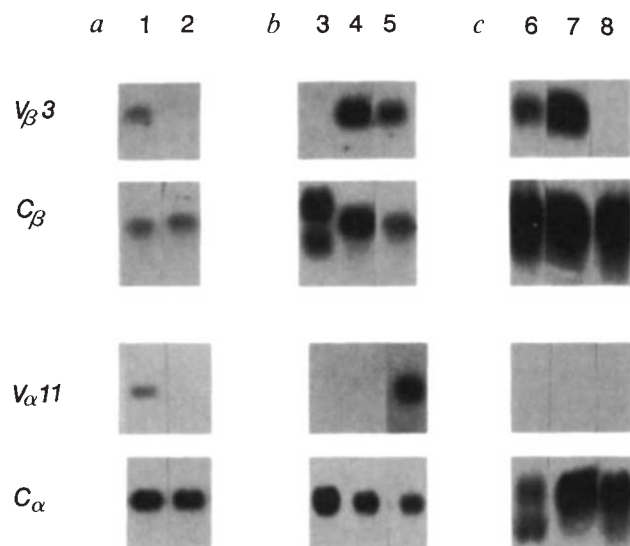
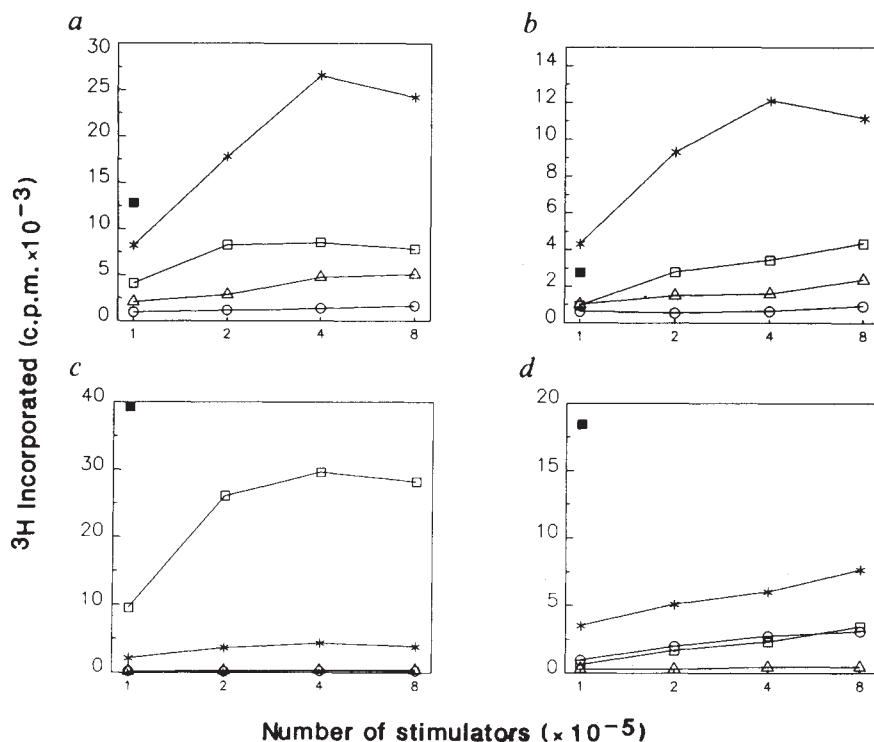


Fig. 2 Northern analysis of V_{α} and V_{β} usage in antigen-specific T-cell lines and clones. a, Lane 1, B10.BR anti-cytochrome *c* T-cell line; lane 2, B10.BR anti-PPD T-cell line. b, Lane 3, BCAC3, an Mls^a-specific T-cell clone⁷; lane 4, BCAC5, an Mls^c-specific T-cell clone⁷; lane 5, 1C5, a cytochrome *c*-specific T-cell clone⁵. c, Lane 6, BC3B13, an Mls^c-specific T-cell clone⁸; lane 7, BC3C13, an Mls^c-specific T-cell clone⁸; lane 8, an Mls^a-specific T-cell clone⁸. The Mls^c-specific clones used were shown to be independent (non-identical) by analysis of their independent derivation and/or fine specificity (ref. 8 and data not shown).

Methods. Total RNA was extracted from T-cell lines or clones by a guanidinium isothiocyanate/caesium chloride gradient method¹⁷. RNA (10 μ g per lane except lane 7 which had 15 μ g per lane) was separated by electrophoresis in an 0.8% morpholinopropanesulphonic acid/formaldehyde gel, transferred to Genescreen nylon membranes (NEN) and cross-linked using ultraviolet light. $V_{\beta}3$, C_{β} , $V_{\alpha}11$, C_{α} (refs 3-5), single-copy specific probes, were labelled with ³²P dCTP by nick translation, denatured and hybridized to the filters at 42 °C overnight in 10% dextran sulphate, 4 $\times$ SSC, 40% formamide, 5 $\times$ Denhardt's solution, 20 μ g ml⁻¹ fish sperm DNA and 1% SDS. Filters were washed twice at 55 °C in 0.1% SSC and 0.05% SDS for 30 min and autoradiographed for 12-48 h at -70 °C with intensifying screens. Filters were first hybridized to the variable region probes and were then stripped at 80 °C for 5 min in 0.1% SSC, 0.05% SDS before rehybridization of the same filter with the corresponding TCR constant-region under similar conditions.

particular V_{β} gene segments in the T-cell population can therefore account for the high precursor frequency of T cells specific for Mls^a or Mls^c. Expression of a selected V_{β} is sufficient to confer specificity to Mls^a ($V_{\beta}8.1$ or $V_{\beta}6$, refs 9 and 10), to Mls^c ($V_{\beta}3$) or to $E_{\alpha}E_{\beta}$ ($V_{\beta}17a$; ref. 12) determinants, representing the three sets of gene products that activate high proportions of T cells. These results contrast with the apparent involvement of multiple α - and β -chain segments in determining T-cell specificity for conventional peptide antigens in association with self-MHC products^{3,13}.

If the correlation between Mls^c reactivity and $V_{\beta}3$ expression is strong enough, strains that express the Mls^c product might be expected to delete cells expressing $V_{\beta}3$ because these cells would otherwise be autoreactive. Recently, T cells expressing

$V_{\beta}3$ have been found to be deleted in several Mls^c strains and their F₁ hybrids^{14,15}. Analogous deletions of $V_{\beta}8.1$ and $V_{\beta}6$ in Mls^a mice and of $V_{\beta}17a$ in mice expressing $E_{\alpha}E_{\beta}$ have been described and were attributed to maintenance of self-tolerance^{9,10,12}. The evidence provided here that the T-cell repertoire used in Mls^b strains for responses to cytochrome *c* is highly cross-reactive to Mls^c provides a novel model system in which to evaluate further the effect of tolerance to self-Mls^c determinants on the response to foreign antigen.

The authors thank Dr L. A. Matis for providing T-cell clones and for constructive discussions and Dr F. D. Finkelman for goat anti-mouse antiserum. We also thank Drs R. N. Germain, R. H. Schwartz and A. Singer for critical reviews of this manuscript.

Received 16 June; accepted 25 August 1988.

- Wilson, D. B., Blyth, J. L. & Nowell, P. C. *J. exp. Med.* **128**, 1157-1181 (1968).
- Janeway, C. A. Jr, Lerner, E. A., Jason, J. M. & Jones, B. *Immunogenetics* **5**, 357-371 (1977).
- Fink, P. J., Matis, L. A., McElligott, D. L., Bookman, M. & Hedrick, S. M. *Nature* **321**, 219-226 (1986).
- Sorger, S. B., Hedrick, S. M., Fink, P. J., Bookman, M. A. & Matis, L. A. *J. exp. Med.* **165**, 279-301 (1987).
- Matis, L. A., Sorger, S. B., McElligott, D. L., Fink, P. J. & Hedrick, S. M. *Cell* **51**, 59-69 (1987).
- Festenstein, H. *Transplant Rev.* **15**, 62-88 (1987).
- Abe, R., Ryan, J. J. & Hodes, R. J. *J. exp. Med.* **166**, 1113-1129 (1987).
- Abe, R., Ryan, J. J., Finkelman, F. D. & Hodes, R. J. *J. Immun.* **138**, 373-379 (1987).
- Kappler, J. W., Staerz, U., White, J. & Marrack, P. C. *Nature* **332**, 35-40 (1988).
- MacDonald, J. R. *et al. Nature* **332**, 40-45 (1988).
- Abe, R., Ryan, J. J. & Hodes, R. J. *J. exp. Med.* **165**, 1150-1155 (1987).
- Kappler, J. W. *et al. Cell* **49**, 263-271 (1987).
- Saito, T. & Germain, R. N. *Immunol. Rev.* **101**, 81-114 (1988).
- Fry, A. M. & Matis, L. A. *Nature* **335**, 830-832 (1988).
- Pullen, A. M., Marrack, P. & Kappler, J. W. *Nature* **335**, 796-801 (1988).
- Ryan, J. J., Mond, J. J., Finkelman, F. D. & Scher, I. *J. Immun.* **130**, 2534-2541 (1987).
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294-5303 (1979).

Self-tolerance alters T-cell receptor expression in an antigen-specific MHC restricted immune response

Alicia M. Fry & Louis A. Matis

Molecular Immunology Laboratory, Division of Biochemistry and Biophysics, Center for Biologics Evaluation and Research, Food and Drug Administration, Bethesda, Maryland 20892, USA

The influence of major histocompatibility complex (MHC) gene products on the T-lymphocyte $\alpha\beta$ receptor (TCR) repertoire is well documented¹⁻⁴, but how specificity is also generated for a diverse array of foreign peptide antigens is unknown. One proposed mechanism is that the TCR repertoire is selected by the recognition of processed self-antigens bound to MHC molecules^{5,6}. Here, we examine the influence of non-MHC-encoded self-antigens on the TCR repertoire expressed in an antigen-specific immune response. Most pigeon cytochrome *c*-specific, $E^k E^k$ (E^k) Ia-restricted T cells from B10.A mice express a product of the $V_{\alpha}11$ gene family in association with a $V_{\beta}3$ gene-encoded protein⁷⁻¹⁰. We therefore examined $V_{\alpha}11$ and $V_{\beta}3$ gene expression in cytochrome *c*-specific T-cell lines derived from various mouse strains with different non-MHC genetic backgrounds. T cells from several strains failed to express any $V_{\beta}3$ due to tolerance induced by MHC-encoded self-antigens^{11,12}. Variable levels of $V_{\alpha}11$ messenger RNA (mRNA) were expressed by antigen-specific T cells from all the strains. In one strain $V_{\beta}3$ was expressed in the relative absence of $V_{\alpha}11$. These results directly demonstrate that self-tolerance alters TCR gene usage in the immune response to a foreign antigen, and indicate that TCR V_{α} and V_{β} proteins may, in part, be independently selected.

Antigen-specific E^k -restricted T-cell lines were established *in vitro* following immunization of mice with either intact pigeon cytochrome *c* (Fig. 1a) or the carboxy-terminal fragment 81-104 (Fig. 1b). RNA was extracted from each of these lines and probed for $V_{\beta}3$ and $V_{\alpha}11$ gene expression. $V_{\beta}3$ mRNA was expressed by cytochrome *c*-specific T-cell lines derived from B10.BR, AKR/J, C57BR and B10.A, but not CBA/J, C3H/HeJ, A.TL, or A/J mice (Fig. 1). Significantly, the [B10.A ($V_{\beta}3^+$) $\times$ A/J($V_{\beta}3^-$)]F₁ T-cell line failed to express any $V_{\beta}3$ mRNA (Fig. 1b).

Analysis of alternative V_{β} gene usage by the $V_{\beta}3^-$ T cells revealed variable expression by all the lines of $V_{\beta}1$ mRNA, which is frequently expressed by *I-E* restricted cytochrome *c*-specific T cells^{8,10,13,14} and very little or no expression of $V_{\beta}5$ mRNA, which has not been reported in any cytochrome *c*-specific T cells (Fig. 1).

Variable levels of $V_{\alpha}11$ mRNA were found in all the T-cell

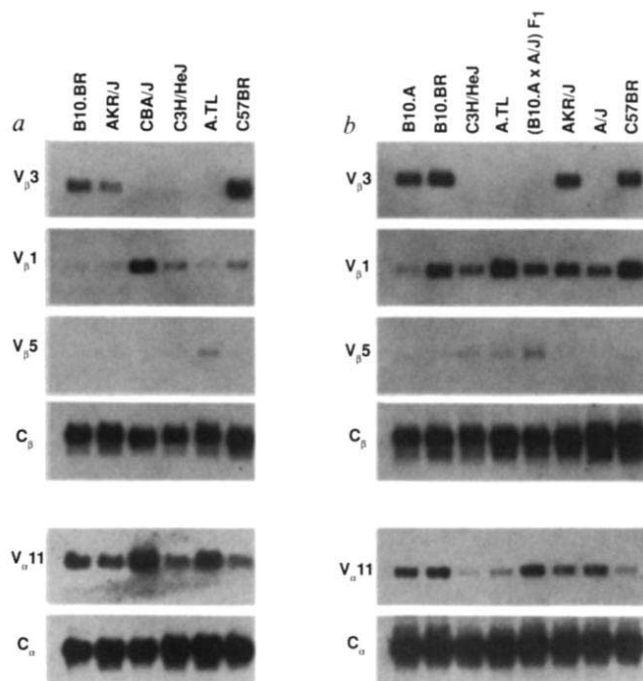


Fig. 1 TCR β and α mRNA expression in pigeon cytochrome *c*-specific E^k -restricted T-cell lines. RNA was prepared from all the lines after four or five cycles of antigen stimulation *in vitro*, when the lines consist almost entirely of antigen-specific T cells (L. Matis, unpublished data). The T-cell lines were derived from mice immunized with either whole pigeon cytochrome *c* 1-104 (a) or pigeon cytochrome *c* fragment 81-104 (b). The variable (V_{β} and V_{α}) and constant (C_{β} and C_{α}) region probes have been previously described^{7,9,10}. The $V_{\beta}5$ probe was provided by Dr Dennis Loh (Washington University)²². The strain of origin of the cytochrome *c*-specific T-cell line from which the RNA was prepared is shown above each lane. Equivalent levels of TCR β mRNA in all the lines were shown by hybridization to the TCR constant region (C_{β}) probe. Southern hybridization confirmed that the $V_{\beta}3$ and $V_{\alpha}11$ gene families were present in germ line DNA of all the strains (data not shown).

Methods. Total cellular RNA was prepared following extraction in guanidinium isothiocyanate²³. RNA (10 μ g per lane) was subjected to electrophoresis on 1% agarose/0.7% formaldehyde gels at 100 V for 4 h before transfer to nitrocellulose filters. Preparation of the probes, hybridizations and washes were performed as previously described^{9,10}. The data represent overnight exposures to X-ray film with the exception of $V_{\alpha}11$ in a, which represents a six-day exposure.

lines (Fig. 1). However, the relative amount of $V_{\alpha}11$ mRNA expressed ($V_{\alpha}11/C_{\alpha}$) did not correlate with $V_{\beta}3$. For example, $V_{\alpha}11$ expression was low in the C3H/HeJ($V_{\beta}3^-$) and C57BR($V_{\beta}3^+$) lines, but high in the $V_{\beta}3^-$ CBA/J, A.TL, (B10.A $\times$ A/J)F₁ and A/J lines.

The strain-related $V_{\beta}3$ deletion from the cytochrome *c*-specific TCR repertoire was explored further by examining levels of V_{β} RNA in poly(A)⁺ mRNA prepared from concanavalin A-stimulated splenic T-cell blasts (Fig. 2). $V_{\beta}3$ mRNA was expressed by the B10.A and AKR/J T cells, but not the C3H/HeJ, A/J, or (B10.A $\times$ A/J)F₁ populations. In contrast, approximately equal amounts of $V_{\beta}1$ and $V_{\beta}5$ mRNA were expressed by all the lectin-stimulated T cells. Thus, $V_{\beta}3$ is not expressed in any peripheral T cells of certain mouse strains.

All the $V_{\beta}3^-$ strains shown here express MHC^c (refs 15, 16). The association of $V_{\beta}3$ expression with MHC^c reactivity^{11,12}, the failure of a ($V_{\beta}3^+ \times V_{\beta}3^-$)F₁ hybrid to express $V_{\beta}3$ (Figs 1 and 2) and the detection of $V_{\beta}3$ mRNA in adult thymocytes from MHC^c-positive mice (data not shown), indicate that self-tolerance

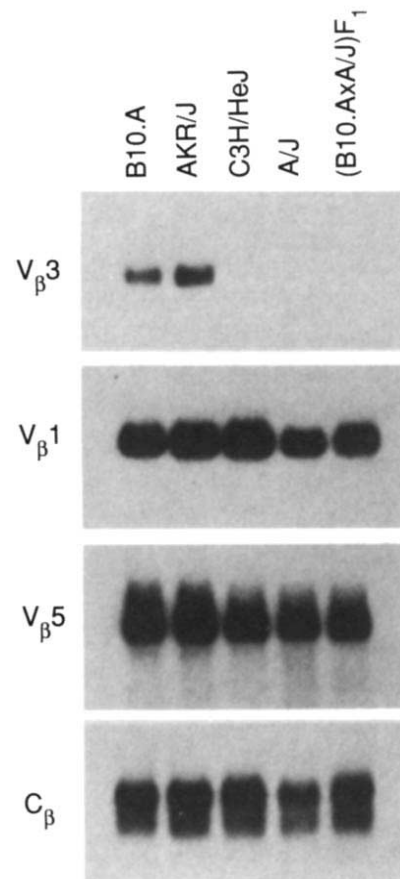
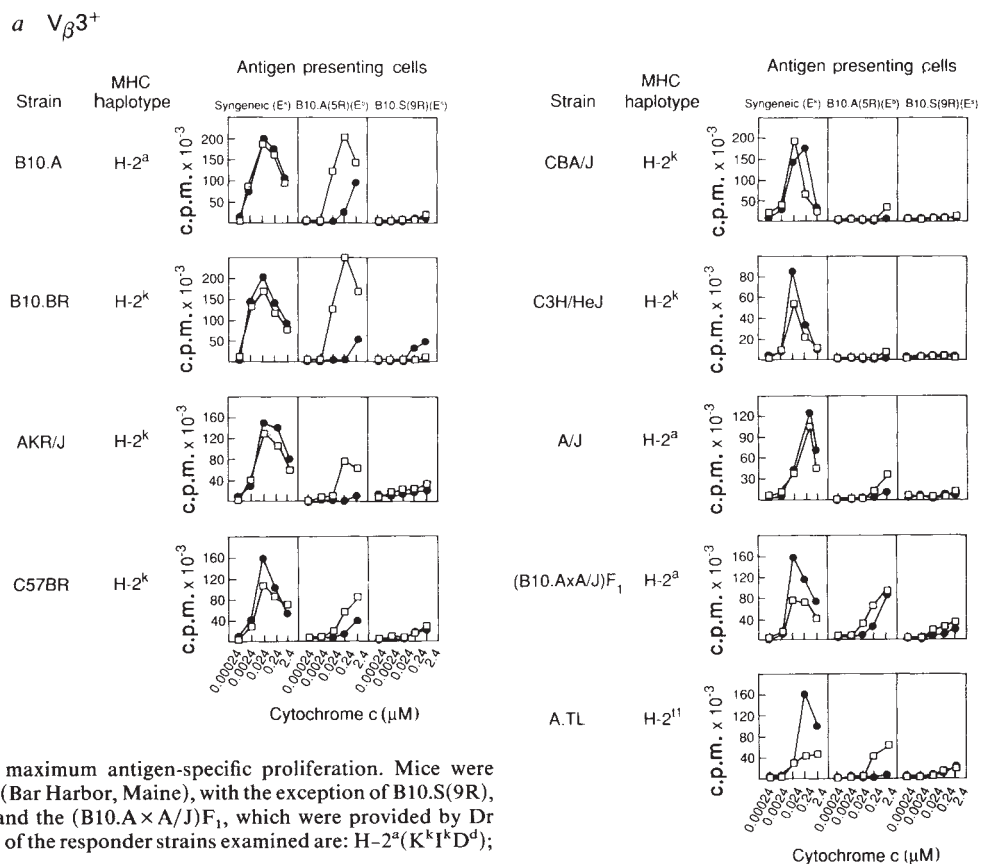


Fig. 2 Expression of $V_{\beta}3$, $V_{\beta}1$, $V_{\beta}5$ and C_{β} genes in poly(A)⁺ mRNA prepared from lectin-stimulated splenic T-cell blasts from B10.A, AKR/J, C3H/HeJ, A/J and (B10.A x A/J)F₁ mice.

Methods. Spleens were collected from non-immunized mice. Single-cell suspensions were prepared, the red blood cells lysed, and 2×10^6 cells ml⁻¹ were cultured in complete medium^{9,10} with $5 \mu\text{g ml}^{-1}$ of concanavalin A (Sigma). After 48 h, the cells were collected and total cellular RNA prepared. Poly(A)⁺ RNA was isolated by oligo(dT)-cellulose chromatography²⁴ (Type 7, Pharmacia). Electrophoresis was performed on $8 \mu\text{g}$ per lane of poly(A)⁺ RNA, with the exception of the first two lanes of the filter hybridized to $V_{\beta}5$ only and containing $16 \mu\text{g}$ of poly(A)⁺ RNA each. The V_{β} and C_{β} probes were the same ones used to examine mRNA from the cytochrome *c*-specific T-cell lines. The data represent overnight exposures to X-ray film.

Fig. 3 Comparison of the antigen and MHC specificities of E^k-restricted pigeon cytochrome *c*-specific T-cell lines derived from $V_{\beta}3$ expressing (a) and $V_{\beta}3$ non-expressing (b) mouse strains. The proliferative responses to various concentrations of cytochrome *c* 88-104 (●—●) and 88-103 (□—□) peptides were measured in the presence of irradiated (3,300R) APCs from syngeneic ($E^k_a E^k_b : E^k$), or B10.A(5R) ($E^k_a E^k_b : E^b$), and B10.S(9R) ($E^k_a E^k_b : E^s$) mice. The 88-104 and 88-103 synthetic peptides (the gift of Dr Stephen M. Hedrick, University of California) elicit response patterns similar to the native 81-104 pigeon cytochrome *c* and 81-103 moth cytochrome *c* peptides, respectively^{13,25}. The sequence of 88-103 is identical to 88-104, except that an alanine has been deleted at residue 103, retaining the carboxy-terminal lysine. The failure of B10.A(4R) APC (A^k, E-negative) to present either peptide established E^k as the restricting Ia molecule in all cases (data not shown). The c.p.m. shown represent the arithmetic mean of duplicate cultures. Background c.p.m. in all cases were <10% of the maximum antigen-specific proliferation. Mice were purchased from The Jackson Laboratory (Bar Harbor, Maine), with the exception of B10.S(9R), which were bred in our animal colony and the (B10.A x A/J)F₁, which were provided by Dr David Sachs, NCI. The MHC haplotypes of the responder strains examined are: H-2^a(K^kI^kD^d); H-2^k(K^kI^kD^k); H-2^u(K^sI^kD^d).

Methods. T-cell lines were established following immunization with either $100 \mu\text{g}$ of intact pigeon cytochrome *c* 1-104 or 20 nmol of the carboxy terminal 81-104 fragment (kindly provided by Dr Ronald H. Schwartz, NIH). The propagation of the antigen-specific proliferative cell lines has been previously described^{9,10}. The proliferation assays were performed exactly as described previously^{9,10} on T-cell lines following four or five cycles of *in vitro* antigen stimulation.



to Mls^c-encoded antigens eliminates peripheral V_β3-expressing T cells. This is similar to the reports of Mls^a reactivity mediated by the V_β8.1 and V_β6 genes^{6,17}.

In addition, our data have particular implications for TCR selection in antigen-specific immune responses. For example, the expression of V_α11 in cytochrome *c*-specific T-cell lines without V_β3 and conversely the expression of V_β3 in the C57BR T-cell line in the relative absence of V_α11 (both of the cytochrome *c*-specific C57BR T-cell clones examined express V_β3 but not V_α11; data not shown) suggest that, to some extent, the TCR α- and β-V gene elements are independently selected. This is consistent with a recently proposed structural model of T-cell recognition in which the TCR V_α and V_β proteins contact distinct sites on the MHC molecule/peptide ligand through their complementarity determining regions, CDR1 and CDR2¹⁸.

Finally, the skewing of the cytochrome *c*-specific TCR repertoire alters the fine specificity of the response. The data in Fig. 3 show that the T-cell lines from all the strains examined proliferate in response to both cytochrome *c* peptides 88–104 and 88–103 presented by syngeneic (E^k) antigen-presenting cells (APC). The

B10.A and B10.BR lines manifest the dominant response pattern of cytochrome *c*-specific T cells reported previously^{9,19} (proliferation to both peptides presented by E^k APC, preferential response to 88–103 with E^b APC and very low response to E^s APC; Fig. 3a). Analyses of TCR usage in the cytochrome *c* response have shown a direct correlation between the capacity to recognize the 88–103 peptide presented by E^b (B10.A(5R)) APC and V_β3 expression^{9,13,20}. In this context, examination of the response patterns of the V_β3⁺ cytochrome *c*-specific T-cell lines revealed that three of the five (CBA/J, C3H/HeJ and A/J) respond poorly to E^b/88–103 (Fig. 3b). These results raise the possibility that E^b haplotype, Mls^c-positive mice would be low responders following immunization with cytochrome *c* peptide 88–103. This would represent a rare example of an immune response gene defect resulting from a hole in the T-cell repertoire²¹.

We thank Drs Ronald N. Germain, Ronald H. Schwartz and Alfred Singer for critical review of this manuscript. We thank Ms Ellen Kirshbaum for preparation of the manuscript. A.M.F. is a Research Scholar of the Howard Hughes Medical Institute.

- Schwartz, R. H. *Adv. Immun.* **38**, 31–201 (1986).
- Shevach, E. M. & Rosenthal, A. S. *J. exp. Med.* **138**, 1213–1229 (1973).
- Zinkernagel, R. M. *et al. J. exp. Med.* **147**, 882–896 (1978).
- Kappler, J. W., Roehm, N. & Marrack, P. *Cell* **49**, 273–280 (1987).
- Singer, A., Munitz, T. L. & Gress, R. E. *Transplant Proc.* **19**, 107–110 (1987).
- Kappler, J. W., Staerz, U., White, J. & Marrack, P. C. *Nature* **332**, 35–40 (1988).
- Fink, P. J., Matis, L. A., McElligott, D. L., Bookman, M. A. & Hedrick, S. M. *Nature* **321**, 219–226 (1986).
- Winoto, A. *et al. Nature* **324**, 679–682 (1986).
- Sorger, S. B., Hedrick, S. M., Fink, P. J., Bookman, M. A. & Matis, L. A. *J. exp. Med.* **165**, 279–301 (1987).
- Matis, L. A., Sorger, S. B., McElligott, D. L., Fink, P. J. & Hedrick, S. M. *Cell* **51**, 59–69 (1987).
- Abe, R., Vacchio, M. S., Fox, B. & Hodes, R. J. *Nature* **335**, 827–830 (1988).
- Pullen, A. M., Marrack, P. & Kappler, J. W. *Nature* **335**, 796–801 (1988).
- McElligott, D. L., Sorger, S. B., Matis, L. A. & Hedrick, S. M. *J. Immun.* **140**, 4123–4131 (1988).
- Hedrick, S. M. *et al. Science* **239**, 1541–1544 (1988).
- Abe, R., Ryan, J. J., Finkelman, F. D. & Hodes, R. J. *J. Immun.* **138**, 373–379 (1987).
- Abe, R., Ryan, J. J. & Hodes, R. J. *J. exp. Med.* **166**, 1113–1129 (1987).
- MacDonald, H. R. *et al. Nature* **332**, 40–45 (1988).
- Davis, M. M. & Björkman, P. J. *Nature* **334**, 395–402 (1988).
- Heber-Katz, E. *et al. J. exp. Med.* **155**, 1086–1099 (1982).
- Saito, T. & Germain, R. N. *Nature* **329**, 256–259 (1987).
- Ogasawara, K., Maloy, W. L. & Schwartz, R. H. *Nature* **325**, 450–452 (1987).
- Behlke, M. A. *et al. Science* **229**, 566–570 (1985).
- Chirgwin, J. M., Przybyla, A. E., Macdonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294–5303 (1979).
- Aviv, H. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **69**, 1408–1412 (1972).
- Schwartz, R. H., Fox, B. S., Fruga, E., Chen, C. & Singh, B. J. *Immun.* **135**, 2598–2608 (1985).

Glucocorticoid-dependent oncogenic transformation by type 16 but not type 11 human papilloma virus DNA

Mary M. Pater, Glenys A. Hughes, David E. Hyslop, Harikrishna Nakshatri & Alan Pater

Basic Medical Sciences, Faculty of Medicine, Memorial University of Newfoundland, St John's, Newfoundland A1B 3V6, Canada

Squamous cell carcinoma of the uterine cervix is one of the most common cancers among women. Correlation between human papilloma virus (HPV) infection of the uterine cervix and the development of cervical neoplasia has been established^{1–6}. More recent studies have shown the presence and expression of integrated HPV types 16 and 18 DNA sequences in 70–80% of cervical tumours and tumour cell lines^{7–14}. It has been suggested that, in addition to HPVs, other agents such as hormones and tobacco products act as cofactors in cervical neoplasia (for review see ref. 15). The presence and expression of a glucocorticoid-responsive element¹⁶ in HPV-16 has been reported¹⁷. Here we provide evidence for the oncogenic transformation of primary cells with a combination of HPV-16 DNA, but not HPV-11 DNA, and the activated form of the human Ha-ras oncogene only in the presence of the glucocorticoid hormone dexamethasone.

Primary baby rat kidney (BRK) cells were cultured from 8-day-old inbred Fischer rats and grown to subconfluency in Dulbecco's modified Eagle's medium plus 10% fetal bovine serum. HPV-16 and HPV-11 DNAs cloned at their unique BamHI site (in the late region) into pBR322 and pUC19 respectively were cotransfected with the activated form of the human Ha-ras oncogene (EJ c-Ha-ras-1) (10 µg plasmid) into BRK cells¹⁸. The cells were then incubated to give colony formation

in media containing 2% fetal bovine serum either with or without 10^{−6} M dexamethasone; media were changed weekly. Colonies were detected after 2–3 weeks of incubation. No colony formation was observed for HPV or Ha-ras DNAs alone, with or without dexamethasone. As indicated in Fig. 1, the BRK cells transfected with a combination of HPV-16 DNA and Ha-ras oncogene and grown in the presence of dexamethasone could form colonies. No transformation was achieved with HPV-11 DNA and Ha-ras either in the presence or absence of dexamethasone. We tested for HPV DNA in two monoclonal and two polyclonal transformed cell lines. The colonies were expanded and high molecular weight DNA isolated and ana-

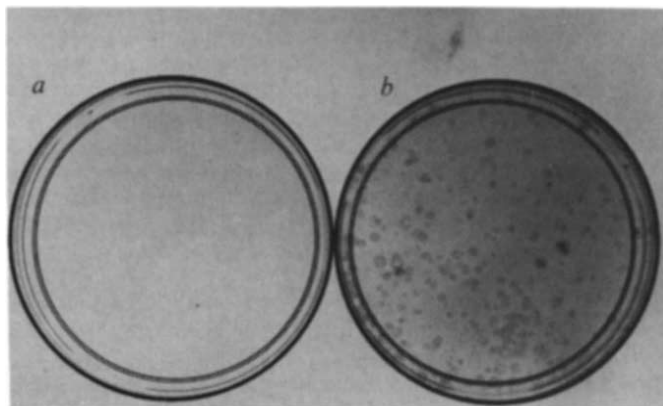
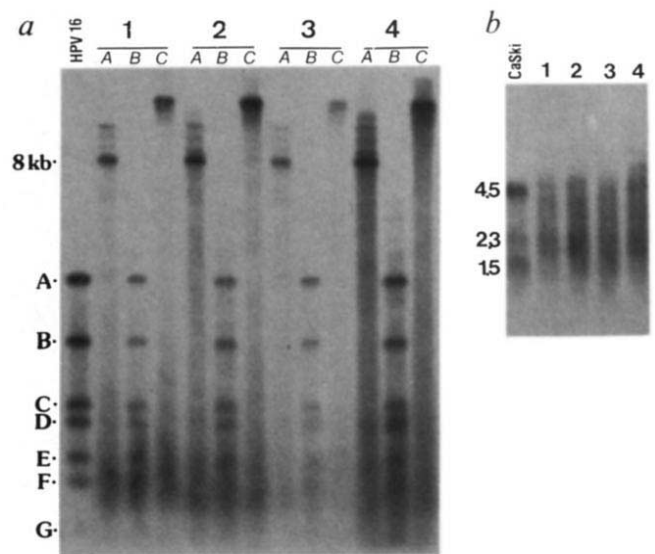


Fig. 1 Transformation of primary BRK cells. BRK cells were transfected with a combination of HPV-16 DNA and Ha-ras oncogene and then incubated *a*, in the absence, or *b*, in the presence of 10^{−6} M dexamethasone. Colonies were detected by staining of the plates after three weeks of incubation.

Fig. 2 Southern and Northern blot analysis of DNA and RNA from transformed colonies. *a*, High molecular weight DNA (10 µg per well) was cleaved with either a non-cut (*Xba*I), single-cut (*Bam*HI) or a multi-cut (*Bam*HI/*Pst*I) restriction enzyme. *Bam*HI/*Pst*I-cleaved HPV-16 DNA (100 pg per well) was used as marker. Samples 1 and 3 are DNA from two monoclonal cell lines and samples 2 and 4 are DNA from two polyclonal cell lines. DNA in lanes A, B and C had been cleaved with *Bam*HI, *Bam*HI/*Pst*I, and *Xba*I restriction enzymes respectively. *b*, Poly(A)⁺ RNA was isolated as described³⁸, with the exception that polysomal RNA was pelleted by centrifugation at 100,000 g for one hour and poly(A)⁺ RNA was then isolated by chromatography on oligo(dT) columns. Northern-blot analysis of RNA (5 µg per well) using formaldehyde gels was as described³⁹. Hybridization was done using total HPV-16 DNA as probe. RNA (2 µg per well) from a cervical carcinoma cell line containing HPV-16 DNA (CaSki)¹⁰ was used as marker. Lanes 1 and 4 are RNA from the two monoclonal cell lines of Fig. 2*a* and lanes 2 and 3 are RNA from the two polyclonal cell lines, respectively.



lysed by Southern blotting¹⁹. The results of digestion of cellular DNA with *Xba*I, which does not cleave HPV-16 plasmid DNA, show that viral DNA is present in the high molecular weight DNA bands, indicating integration into cellular DNA (Fig. 2*a*). The results of digestion with *Bam*HI show the presence of full-length 8-kb HPV DNA in these cells, indicating tandemly repeated copies. Also, all the *Pst*I fragments of HPV-16 DNA are present in these cells, demonstrating the integration of the intact viral genome.

Northern-blot hybridization of cytoplasmic poly(A)⁺ RNA was undertaken to determine whether HPV DNA was expressed in the transformed colonies. All the colonies contained RNA that hybridized with HPV-16 DNA. Several RNA species and a diffuse background can be seen (Fig. 2*b*). It is unlikely that this broad diffuse background is due to degradation of RNA during either extraction or blotting, because both 18S and 28S ribosomal RNA bands appear intact after transfer. Moreover, RNA isolated under identical conditions from a cervical carcinoma cell line containing HPV-16 DNA¹⁰ contains distinct bands hybridizing with HPV-16 DNA. We are investigating whether sequences near the 3' end of the early region affect the stability of viral RNA. HPV-16 or HPV-18 DNA are present and expressed in most tumours and tumour cells of the cervix. Interestingly the majority of tumour cells contain and express the E6-E7 region of viral DNA^{8,11-14}. The transforming activity of the E7 gene of HPV-16 and HPV-18 has also been reported²⁰⁻²³. It was therefore of interest to determine whether this region of viral DNA is expressed in our transformed colonies. Cytoplasmic poly(A)⁺ RNA was analysed by Northern-blot hybridization using the E6-E7 region of viral DNA as probe. All the cell lines contain a major RNA species of ~2 kilobases (kb) which hybridizes with the E6-E7 probe (data not shown).

As stated earlier, transformation of BRK cells with HPV-16 in combination with the *ras* oncogene is dependent on the presence of dexamethasone in the media. We tested whether growth of the transformed colonies in the presence and absence of dexamethasone was also affected. Cell growth was 5-6-fold more in the presence of the hormone (data not shown). Next we assayed for anchorage-independent growth by colony formation in soft agar. Figure 3 indicates that the transformed colonies are capable of anchorage-independent growth. Note also that the number of colonies in agar is much higher when the cells are grown in the presence of the hormone. These cells induced tumours in all suckling rats subcutaneously injected with 5×10^6 cells per animal. These tumours progressed until animals had to be killed two months after injection. No difference in tumour

size was observed when the cells were co-injected with dexamethasone in a microcrystalline form.

Data presented in this report show that transformation of primary BRK cells can be achieved with a combination of Ha-*ras* oncogene and HPV-16 DNA only in the presence of dexamethasone. No transformation is achieved when HPV-11 is used in combination with Ha-*ras* oncogene either in the presence or the absence of the hormone. As we show below, the noncoding region of both HPV-16 (ref. 24) and HPV-11 (ref.

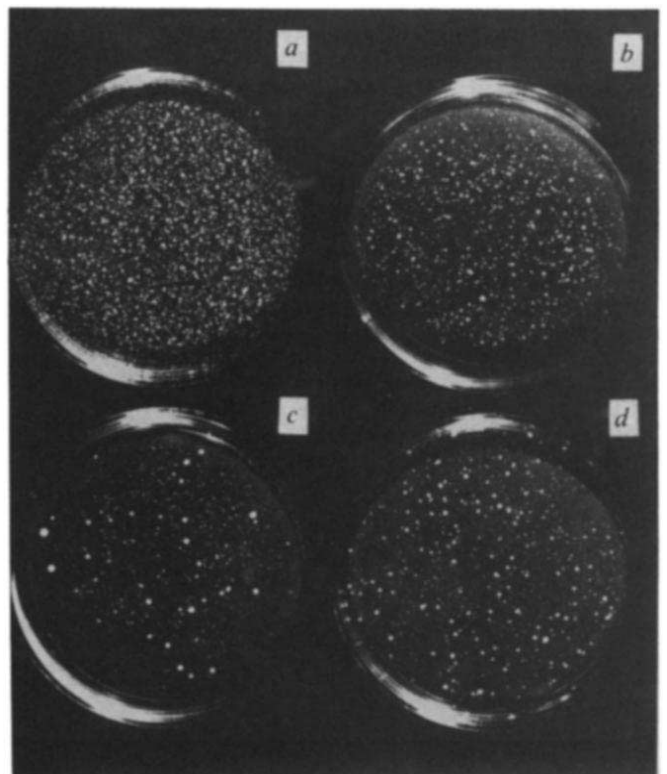


Fig. 3 Colony formation of transformed BRK cells in soft agar. Cells (10^4 cells per plate) were plated in 0.28% agar in the presence and absence of 10^{-6} M dexamethasone. *a* and *b*, Clones 1 and 3 respectively (see Fig. 2*a*) grown in the presence of dexamethasone. *c* and *d*, Clones 1 and 3 respectively (see Fig. 2*a*) grown in the absence of dexamethasone in media-containing agar.

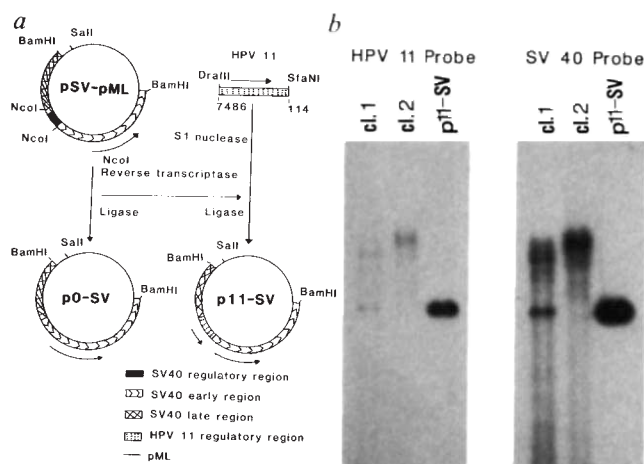


Fig. 4 Analysis of DNA from BRK cells transformed by SV40 early region under the control of HPV-11 non-coding region. *a*, Flow diagram for the construction of p11-SV and p0-SV plasmids. The p11-SV was constructed by cleavage of SV40 DNA (cloned at its unique *Bam*HI site in pML) with *Nco*I to remove the regulatory region together with part of the late region (nucleotides 560-37; ref. 40) and then blunt-ended with reverse transcriptase. The 559-bp *Dra*III-*Sfa*NI (nucleotides 7,486-114) of HPV-11 containing the regulatory region of this virus²⁵ was treated with *S*₁-nuclease and inserted in the correct orientation in the place of the SV40 regulatory region. The *Nco*I-cleaved SV40 DNA lacking its regulatory region was self-ligated to generate the p0-SV to use as a control. Primary BRK cells were transfected with 10 µg of each plasmid together with 10 µg *ras* oncogene and then incubated in the presence or absence of 10⁻⁶ M dexamethasone. An average of 63 and 5 colonies per dish (duplicate plates in two separate experiments) for cells transfected with p11-SV plus *ras* were detected after two to three weeks of incubation in the presence or absence of dexamethasone respectively. No colonies were detected for cells transfected with p0-SV plus *ras* either in the presence or absence of dexamethasone. *b*, High molecular weight DNA from two of the colonies transformed in the presence of dexamethasone was isolated and blotted in duplicate after cleavage with the restriction endonuclease *S*all which has one cleavage site for p11-SV. Note that only clone 1 (cl. 1) has the p11-SV integrated in tandemly repeated copies.

25) contains sequences closely similar to the consensus glucocorticoid-responsive element (GRE) sequence²⁶.

	7,832												7,817
HPV-11	G	G	T	A	C	A	(N ₄)	T	G	T	T	C	A
	.	.	.	.	.	.	.	.	.	.	.	.	.
GRE	G	G	T	A	C	A	(N ₃)	T	G	T	T	C	T
	.	.	.	.	.	.	.	.	.	.	.	.	.
HPV-16	T	G	T	A	C	A	(N ₃)	T	G	T	C	A	T
	7,640												7,654

Moreover, the results of transformation experiments in which the early region of SV40 is placed under the control of the non-coding region of HPV-11 (Fig. 4) clearly show that this sequence responds to glucocorticoid hormones. This strongly suggests that the transforming activity of HPV-16 is due to the glucocorticoid-dependent expression of a gene(s) specifically present in the genome of this virus, but not in that of HPV-11. This suggestion is supported by a recent report²¹ in which HPV-11 DNA expressed under the control of a heterologous promoter lacks the ability to cooperate with *ras* in transformation of BRK cells. As HPV-16 and HPV-18 are associated with most cervical neoplasia, and HPV-6 and HPV-11 generally only with benign lesions of the external genitalia. The *in vitro* studies

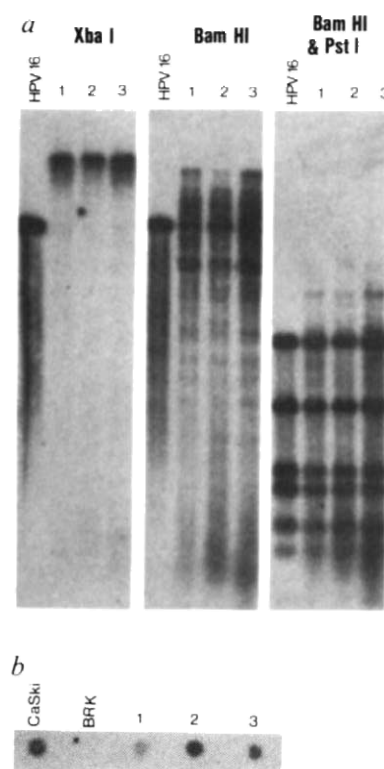


Fig. 5 Analysis of DNA and RNA from BRK cells transformed by a combination of HPV-16 DNA and *ras* oncogene in the presence of 10⁻⁶ M progesterone. *a*, Southern-blot analysis of DNA from three transformed colonies. *b*, Spot-blot hybridization of RNA from the same colonies. Isolation of DNA and RNA and hybridizations were as in Fig. 2. The samples 1, 2 and 3 are DNA or RNA from three monoclonal cell lines. CaSki RNA (2 µg) and BRK RNA (5 µg) were spotted as positive and negative controls respectively.

presented above establish a method to determine the oncogenic potential of HPVs associated with various lesions.

The requirement for the action of the oncogenic HPV-16 genes and the *ras* oncogene for transformation of primary BRK cells is not surprising: multistep transformation and the necessary action of more than one oncogene in the transformation of primary cells have been well documented^{27,28}. Involvement of *ras* in multistep progression of cervical tumours is supported by reports^{29,30} of the amplification of Ha-*ras* oncogene in such tumours.

The same nucleotide sequences that confer responsiveness to glucocorticoids also mediate progesterone induction³¹. Our recent transformation assays have shown colony formation by BRK cells transfected with a combination of HPV-16 DNA and *ras* oncogene in the presence of progesterone. Preliminary analysis of three selected transformed colonies showed the presence and expression of HPV-16 DNA in all the clones (Fig. 5). Moreover, all the colonies induced tumours in suckling rats. The oncogenic potential of progestins, the closely related derivatives of progesterone and the pharmacologically active components of oral contraceptives, remains to be investigated. The urgency of such studies is supported by previous findings that oral contraceptive use can be a potential risk in cervical neoplasia³²⁻³⁷. It is also important to determine whether the incidence of cervical neoplasia in oral contraceptive users is correlated with simultaneous HPV infection in such patients.

We thank T. Marshall for the computer search for GRE homologous sequences, L. Gissmann and H. zur Hausen for HPV-16 and HPV-11 DNA. This work was supported by grants from MRC and NCI of Canada.

Received 20 June; accepted 26 August 1988.

1. Meisels, A. & Fortrin, R. *Acta cytol.* **20**, 505-509 (1976).
2. Puroila, E. & Savia, E. *Acta cytol.* **21**, 26-31 (1977).
3. Lavery, C. R., Russel, P., Hills, E. & Booth, N. *Acta cytol.* **22**, 195-201 (1978).
4. Walker, P. G. *et al. Br. J. Cancer* **48**, 99-101 (1983).
5. Morin, C. & Meisels, A. *Acta cytol.* **24**, 82-84 (1980).
6. Syrjänen, K. *J. Surg. Gynec. Obstet.* **150**, 372-376 (1980).
7. Durst, M., Gissmann, L., Ikenberg, H. & zur Hausen, H. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3812-3815 (1983).
8. Schwarz, E. *et al. Nature* **314**, 111-114 (1985).
9. Yee, C., Krishnan-Hewlett, I., Baker, C. C., Schlegel, R. & Howley, P. M. *Am. J. Path.* **119**, 361-366 (1985).
10. Pater, M. M. & Pater, A. *Virology* **145**, 313-318 (1985).
11. Schneider-Gadicke, A. & Schwarz, E. *EMBO J.* **5**, 2285-2292 (1986).
12. Smolkin, D. & Wettstein, F. O. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4680-4684 (1986).
13. Seedorf, K., Oltersdorf, T., Krammer, G. & Rowekamp, W. *EMBO J.* **6**, 139-144 (1987).
14. Baker, C. C. *et al. J. Virol.* **61**, 962-971 (1987).
15. Peto, R. & zur Hausen, H. *Viral Etiology of Cervical Cancer* Banbury Report 21 (Cold Spring Harbor Laboratory, New York, 1986).
16. Miesfeld, R., Godowski, P. J., Maler, B. A. & Yamamoto, K. R. *Science* **236**, 423-427 (1987).
17. Gloss, B., Bernard, H. U., Seedorf, K. & Klock, G. *EMBO J.* **6**, 3735-3743 (1987).
18. Pater, A. & Pater, M. M. *J. Virol.* **58**, 680-683 (1986).
19. Pater, M. M., Dunne, J., Hogan, G., Ghatage, P. & Pater, A. *Virology* **155**, 13-18 (1986).
20. Phelps, W. C., Lee, C. L., Munger, K. & Howley, P. M. *Cell* **53**, 539-547 (1988).
21. Storey, A. *et al. EMBO J.* **7**, 1815-1820 (1988).
22. Kanda, T., Furono, A. & Yoshiike, K. *J. Virol.* **62**, 610-613 (1988).
23. Watanabe, S. & Yoshiike, K. *Int. J. Cancer* **41**, 896-900 (1988).
24. Seedorf, K., Krammer, G., Durst, M., Suhai, S. & Rowekamp, W. G. *Virology* **145**, 181-185 (1985).
25. Dartmann, K., Schwarz, E., Gissmann, L. & zur Hausen, H. *Virology* **151**, 124-130 (1986).
26. Jantzen, H.-M. *et al. Cell* **49**, 29-38 (1987).
27. Ruley, H. E. *Nature* **304**, 602-606 (1983).
28. Land, H., Parada, L. F. & Weinberg, R. A. *Nature* **304**, 596-602 (1983).
29. Riou, G. E., Barrois, M., Dutronquay, V. & Orth, G. *UCLA Symp. molec. cell. Biol.* **32**, 47-56 (1985).
30. Riou, G. E., Barrois, M., Dutronquay, V. & Orth, G. in *Papilloma Viruses, Molecular and Clinical Aspects* (eds Howley, P. & Broker, T.) 47-56 (Liss, New York, 1985).
31. Strahle, U., Klock, G. & Schutz, G. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7871-7875 (1987).
32. Harris, R. W. C. *et al. Br. J. Cancer* **42**, 359-369 (1980).
33. World Health Organization *Lancet* **ii**, 1207-1208 (1984).
34. Vessey, M. P., Lawless, M., McPherson, K. & Yeates, D. *Lancet* **ii**, 930-938 (1983).
35. Stern, E., Forsythe, A. B., Youkeles, L. & Coffelt, C. F. *Science* **196**, 1460-1462 (1977).
36. Powell, L. C. & Seymour, R. G. *Am. J. Obstet. Gynec.* **110**, 36-41 (1971).
37. Dallenbach-Hellweg, C. *Gebursh. Frauenheilkde* **42**, 249-254 (1982).
38. Pater, M. M., Pater, A., diMayorca, G. & Smiley, J. R. *Virology* **117**, 536-540 (1982).
39. Maniatis, T., Fritsch, E. F. & Sambrook, J. in *Molecular Cloning, a Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
40. Reddy, V. B. *et al. Science* **200**, 494-502 (1978).

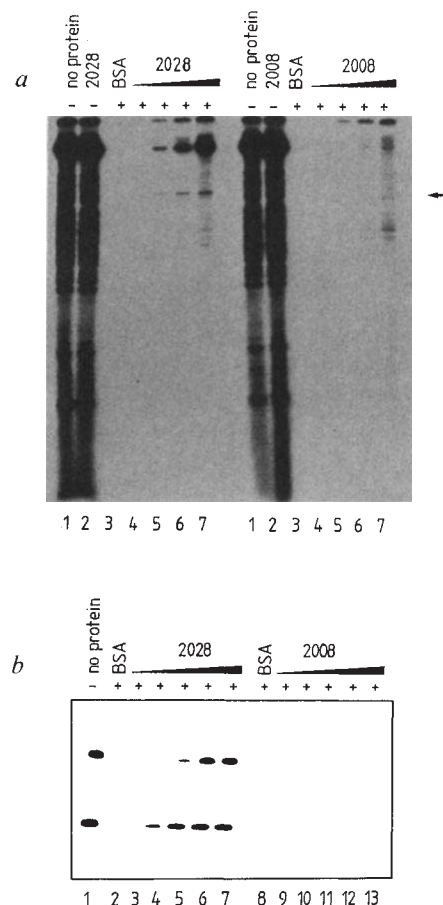


Fig. 1 Specific binding of bacterially expressed v-MYB protein to DNA clones derived from chromatin fragments. DNA from a pool of 25 random plasmid clones of chicken DNA (*a*) or from clone p534, isolated from this pool (*b*), was digested with *EcoRI* and *HindIII*, end-labelled and incubated with full-length (2028) or amino-terminal deleted (2008) v-MYB protein or with bovine serum albumin (BSA). The amount of v-MYB was doubled between adjacent lanes; the highest amount used was 100 ng. Lanes (+), DNA isolated by adsorption to nitrocellulose filters. Lanes (-), total DNA incubated in the absence or presence of protein. DNA samples were analysed on 6% acrylamide gels. The arrow marks the insert of a clone, designated as p534. In *b*, the upper and lower bands represent vector and insert fragments of clone p534 respectively.

Methods. To generate random clones of chicken DNA of uniformly small size, nuclei from chicken myeloblasts transformed by avian myeloblastosis virus were treated with DNase I as described⁴. DNA was purified from the resulting chromatin fragments and cloned into the *SmaI* site of pUC18. Full-length (pVM2028) or amino-terminal deleted (pVM2008) v-MYB protein was expressed in bacteria as described^{5,13}. Proteins were purified as insoluble material (usually 1-2 µg v-MYB per ml bacterial culture)^{5,13} and then solubilized²⁵. The insoluble protein from a 5 ml bacterial culture was dissolved in 200 µl solution containing 10 mM Tris-HCl, pH 7.8, 50 mM NaCl, 1 mM EDTA and 6 M urea, incubated for 10 min at 37 °C and diluted 50-fold into 10 mM Tris-HCl, pH 7.8, 50 mM NaCl, 1 mM EDTA, 250 µg ml⁻¹ BSA at 37 °C, incubated for 30 min at 37 °C and stored on ice. DNA-binding reactions were performed in 100 µl 10 mM Tris-HCl, pH 7.8, 150 mM NaCl, 1 mM EDTA, 1 mM dithioerythritol and 50 µg ml⁻¹ BSA, 2-20 ng end-labelled DNA and 1-20 µl of solubilized v-MYB and were incubated for 30 min at 25 °C. Nitrocellulose filter-binding assays were as described²⁶.

Viral myb oncogene encodes a sequence-specific DNA-binding activity

Horst Biedenkapp, Uwe Borgmeyer*, Albrecht E. Sippel & Karl-Heinz Klempnauer

Zentrum für Molekulare Biologie der Universität Heidelberg, Im Neuenheimer Feld 282, D-6900 Heidelberg, FRG

The retroviral oncogene *v-myb* and its cellular progenitor *c-myb* encode nuclear DNA-binding proteins¹⁻⁵. *Myb* genes have been identified in a broad range of species, including vertebrates⁶⁻⁹, the fruit fly *Drosophila melanogaster*^{10,11} and the plant *Zea mays*¹². The localization of the DNA-binding domain of the v-MYB protein to the highly conserved amino-terminal region^{11,13} suggests that the MYB/DNA interaction is important for MYB function. We show here that v-MYB specifically recognizes the nucleotide sequence pyAAC^G/T^G. So like other nuclear transforming proteins, v-MYB seems to be a member of the class of sequence-specific DNA-binding factors presumably involved in gene regulation.

To investigate the importance of DNA-binding in the function of v-MYB, we determined whether the DNA-binding domain of the protein, which is encoded in the highly conserved amino-terminal third of the amino-acid sequence¹³, recognizes specific nucleotide sequences. Because the target genes for v-MYB which might contain MYB-specific binding sites are unknown, we looked for any sequence-specific DNA-binding indirectly. We used full-length v-MYB protein expressed in bacteria to screen

* Present address: Howard Hughes Medical Institute, Gene Expression Laboratory, The Salk Institute for Biological Studies, PO Box 85800, San Diego, California 92138, USA.

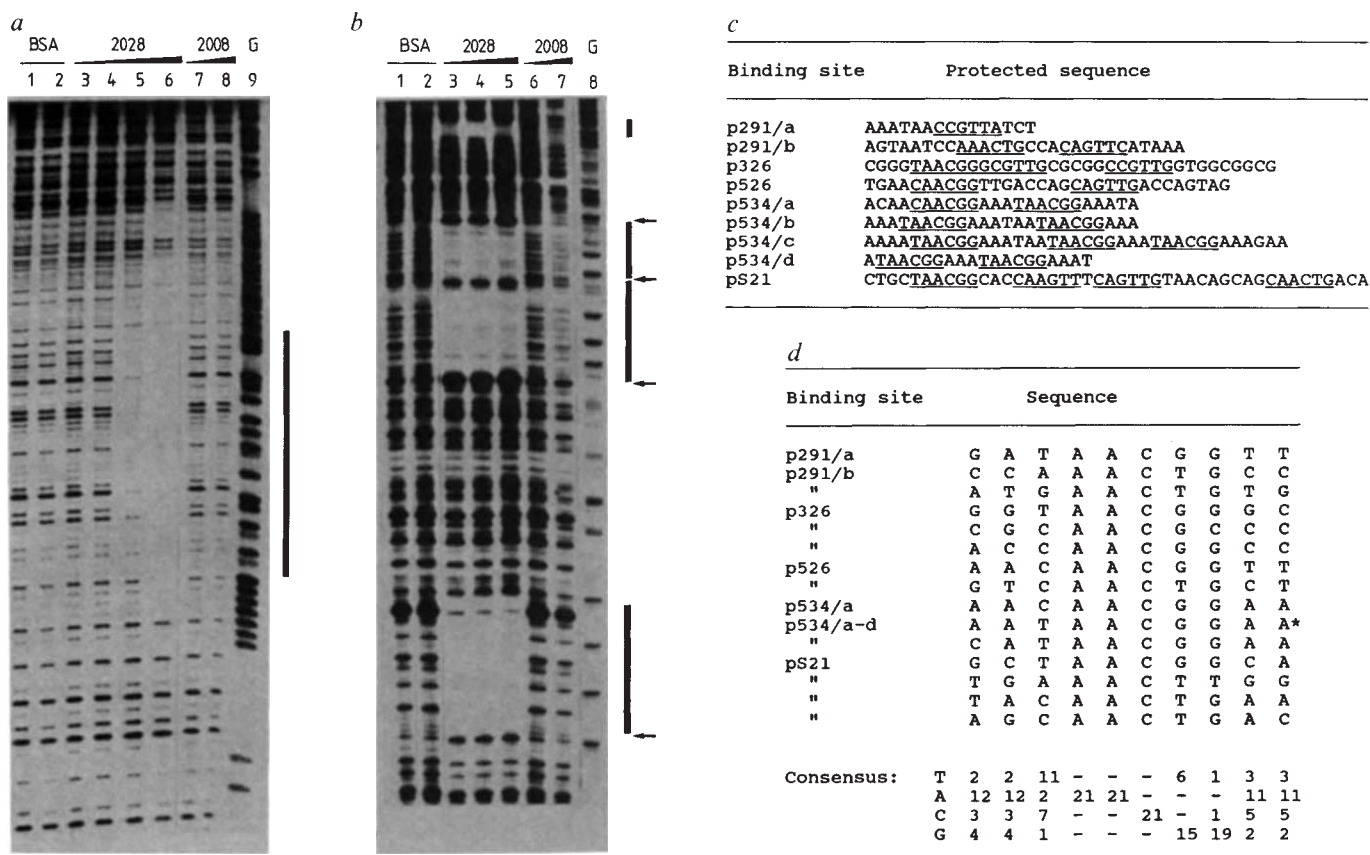


Fig. 2 Identification of MYB-specific binding sites in chromatin-derived DNA clones. *a* and *b*, DNase I footprints of gel-purified full-length v-MYB protein (2028) bound to DNA of clones p326 (*a*) and p534 (*b*). Approximate amounts of v-MYB used were: 80 ng, 160 ng, 240 ng and 320 ng (*a*) and 90 ng, 150 ng and 300 ng (*b*). The other lanes show control reactions containing the same amount of amino-terminal deleted v-MYB (2008) or BSA. Lanes marked G show the products of G-specific sequencing reactions. Bars and arrows mark protections and enhancements of DNase I cutting. *c*, Protected nucleotide sequences identified on six plasmid clones of chicken DNA specifically bound by v-MYB protein. Putative MYB recognition motifs are underlined. Protected sequences p534/a-d are shown on the right in *b* (top to bottom). *d*, Alignment of putative MYB recognition motifs. The sequence marked by an asterisk is present in seven copies in footprint regions p534/b to p534/d.

Methods. DNA-binding reactions, containing 2 to 10 ng of end-labelled DNA-fragment in 100 μ l 10 mM HEPES, pH 8.0, 0.1 mM EDTA, 100 mM NaCl, 2 mM dithiothreitol and 1–20 μ l v-MYB extract or 10 μ l BSA (0.5 mg ml⁻¹), were incubated for 30 min at 37 °C. For footprinting experiments, insoluble v-MYB from a 50 ml culture of bacteria was dissolved in 2 ml 10 mM Tris-HCl, pH 7.8, 50 mM NaCl, 1 mM EDTA and 6 M urea and then dialysed against 10 mM Tris-HCl, pH 7.8, 50 mM NaCl, 1 mM EDTA overnight. The dialysis buffer was prewarmed to 37 °C and cooled to 4 °C at the start of dialysis. When gel-purified v-MYB was used, the protein was electrophoresed in SDS-polyacrylamide gels, recovered from the gel as described^{25,27}, dissolved in urea and dialysed. Footprint reactions were performed as described²⁶.

pools of plasmid clones containing random fragments of chicken DNA for MYB-specific binding sites. The clones were obtained by shotgun-cloning into plasmid pUC18 of chromatin fragments generated by limited DNase I-digestion of chicken myeloblast nuclei⁴. Figure 1*a* shows an example of a nitrocellulose filter-binding assay of a pool of 25 clones, one of which was preferentially retained by v-MYB protein. Figure 1*b* shows the same assay with DNA of an isolated clone, designated as p534. The v-MYB protein specifically bound to the insert fragment of this clone. Any binding to the vector fragment at higher protein concentrations was presumably nonspecific. A partially deleted v-MYB protein lacking the DNA-binding domain did not bind specifically to this clone, indicating that it was v-MYB that was responsible for specific binding. By screening ~1,200 independent clones, we identified 6 clones that bound specifically to v-MYB.

To investigate the sequence requirements for MYB-specific DNA binding, we analysed clones p291, p326, p526, p534 and pS21, all of which bound v-MYB specifically, by DNase I footprinting. In all cases we identified one or more regions protected by v-MYB from nuclease digestion; representative footprinting

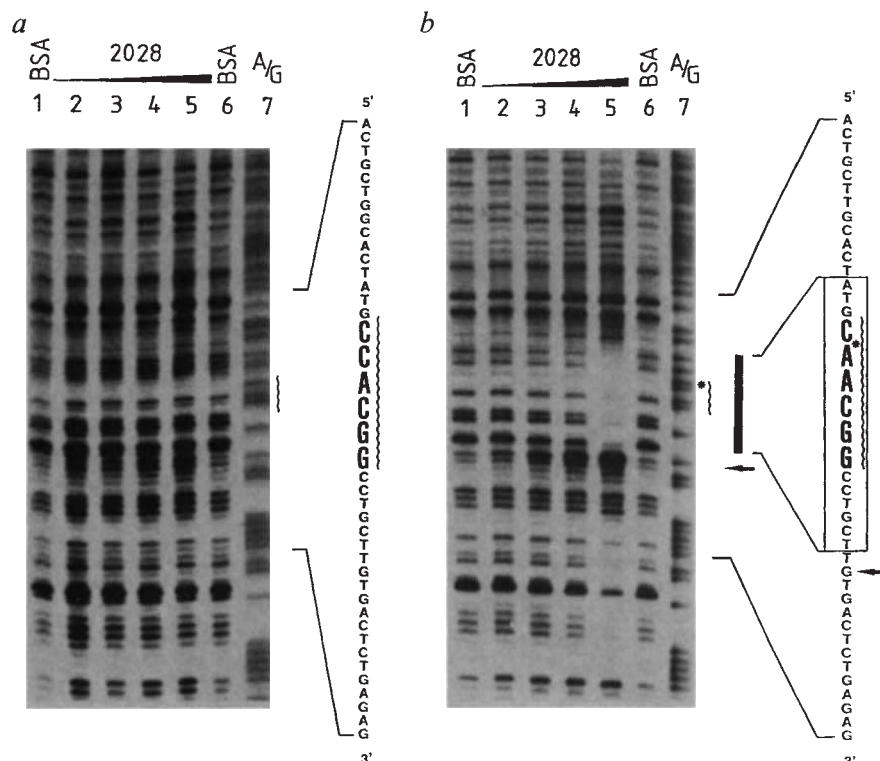
experiments using two of the clones (p326 and p534) are shown in Fig. 2*a* and *b*. Comparison of the protected DNA sequences in each clone revealed a sequence motif, pyAAC^G/T^G, common to all the protected regions (Fig. 2*c* and *d*).

To further test the importance of the pyAAC^G/T^G motif in MYB-specific binding to DNA, we mutated a sequence from an upstream region of the chicken lysozyme gene from CCACGG to CAACGG, thus creating a potential MYB-binding motif. We showed by DNase I footprinting that the newly created binding motif was recognized by v-MYB (Fig. 3). From this result and from the presence of the pyAAC^G/T^G motif in all the footprint regions, we conclude that v-MYB protein is recognizing this motif.

Our findings are in accord with recent demonstrations that the products of other retroviral oncogenes, v-*erb-A* (refs 14 and 15), v-*jun* (refs 16 and 17) and v-*fos* (refs 18 and 19), bind to specific nucleotide sequences, either alone or in conjunction with other proteins. We have noted that v-MYB contains a potential α -helical 'leucine-zipper' sequence²⁰, MIVHQSNI LDNVKNLLEFAETLQLIDSFL (the first residue of this sequence is amino acid 305 in ref. 21), which is related

Fig. 3. The pyAACG/TG-motif is important for MYB-specific DNA-binding. DNase I footprints of gel-purified full-length v-MYB (2028) bound to a sequence derived from the upstream region of the chicken lysozyme gene (panel *a*) or to a mutated version of this sequence (panel *b*). Approximate amounts of v-MYB used were: 50 ng (lane 2), 100 ng (lane 3), 250 ng (lane 4) and 500 ng (lane 5). Other lanes represent control reactions containing BSA instead of v-MYB. Lanes marked by A/G show the products of (A + G)-specific sequencing reactions of the radiolabelled DNA fragments. The position of a point mutation converting the original CCACGG sequence to the putative v-MYB recognition motif CAACGG is marked by an asterisk. The protected region and an enhancement of DNase I cutting are marked by a bar and an arrow respectively.

Methods. Footprinting analysis was carried out as described in Fig. 2, using a *Stu*I/*Bsm*I restriction fragment from about 6 kilobases upstream of the chicken lysozyme gene promoter (see ref. 26 for the exact location of the mutated sequence, designated BS1b).



to sequences found in *myc*-, *fos*- and *jun*-encoded proteins and in the transcription factor C/EBP, and is assumed to mediate protein-protein interaction²⁰. We do not know whether or not these similarities reflect comparable modes of action for v-MYB and these other proteins.

At present we can only speculate on the *in vivo* function of MYB. The 'myb-like' *C1* gene of *Zea mays* controls the expression of genes involved in anthocyanin biosynthesis¹², and *myb* genes could also be regulatory. Interestingly, the *A1* gene of *Zea mays* is a candidate target gene for the *C1* gene product²² and lies close to a cluster of 12 puAACGT sequences, which is related to the MYB-binding motif. As well as indicating that the MYB recognition motif has been highly conserved during evolution, this observation also supports the idea that v-MYB controls gene expression by direct binding to MYB target genes. Our approach could be used to identify MYB target genes in a collection of DNA sequences. Finally, it is interesting that the amino-acid sequences of the MYB DNA-binding domain at its amino terminal¹³ show no significant homology with consensus sequences of either 'helix-turn-helix'- or 'zinc-finger'-type DNA-binding proteins^{23,24}, so MYB could exemplify a novel type of DNA-binding protein.

We thank H. Arnold for technical assistance, J. Paz-Ares, U. Wienand and H. Saedler for communicating results before publication and C. Bonifer, M. Koenen and A. Nordheim for discussion. This work was supported by the Bundesministerium für Forschung und Technologie.

11. Peters, C. W. B., Sippel, A. E., Vingron, M. & Klempnauer, K.-H. *EMBO J.* **6**, 3085-3090 (1987).
12. Paz-Ares, J. *et al.* *EMBO J.* **6**, 3553-3558 (1987).
13. Klempnauer, K.-H. & Sippel, A. E. *EMBO J.* **6**, 2719-2725 (1987).
14. Sap, J. *et al.* *Nature* **324**, 635-640 (1986).
15. Weinberger, C. *et al.* *Nature* **324**, 641-646 (1986).
16. Bohmann, D. *et al.* *Science* **238**, 1386-1392 (1987).
17. Angel, P. *et al.* *Nature* **332**, 166-171 (1988).
18. Rauscher, F. J. *et al.* *Cell* **52**, 471-480 (1988).
19. Rauscher, F. J. *et al.* *Science* **240**, 1010-1016 (1988).
20. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. *Science* **240**, 1759-1764 (1988).
21. Klempnauer, K.-H., Gonda, T. J. & Bishop, J. M. *Cell* **31**, 453-463 (1982).
22. Schwarz-Sommer, Z. *et al.* *EMBO J.* **6**, 287-294 (1987).
23. Klug, A. & Rhoades, D. *Trends biochem. Sci.* **12**, 464-469 (1987).
24. Pabo, C. O. & Sauer, R. T. *A. Rev. Biochem.* **53**, 293-321 (1984).
25. Hager, D. A. & Burgess, R. R. *Analyt. Biochem.* **109**, 76-86 (1980).
26. Borgmeyer, U., Nowock, J. & Sippel, A. E. *Nucleic Acids Res.* **12**, 4295-4311 (1984).
27. Wessel, D. & Flügge, U. I. *Analyt. Biochem.* **138**, 141-143 (1984).

Errata

Crystal structure of *trp* repressor/operator complex at atomic resolution

Z. Otwinowski, R. W. Schevitz, R.-G. Zhang, C. L. Lawson, A. Joachimiak, R. Q. Marmorstein, B. F. Luisi & P. B. Sigler
Nature **335**, 321-329 (1988).

IN this article the present address of one of the authors was given incorrectly. Rong-guang Zhang's correct address is: Chinese Academy of Sciences, Institute of Biochemistry, Shanghai, The People's Republic of China.

Controls on the structure of subducted slabs

Michael Gurnis & Bradford H. Hager
Nature **335**, 317-321 (1988).

THE above title better reflects the meaning of this article, which appeared in the 22 September issue with the word 'of' instead of 'on'.

Received 21 July; accepted 19 September 1988.

1. Klempnauer, K.-H. *et al.* *Cell* **33**, 345-355 (1983).
2. Boyle, W. J. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **81**, 4265-4269 (1984).
3. Moelling, K. *et al.* *Cell* **40**, 983-990 (1985).
4. Klempnauer, K.-H. & Sippel, A. E. *Molec. cell. Biol.* **6**, 62-69 (1986).
5. Klempnauer, K.-H., Bonifer, C. & Sippel, A. E. *EMBO J.* **5**, 1903-1911 (1986).
6. Gonda, T. J. & Bishop, J. M. *J. Virol.* **46**, 212-210 (1983).
7. Gonda, T. J., Gough, N. M., Dunn, A. R. & de Blaquiere, J. *EMBO J.* **4**, 2003-2008 (1985).
8. Rosson, D. & Reddy, E. P. *Nature* **319**, 604-606 (1986).
9. Slamon, D. J. *et al.* *Science* **233**, 347-351 (1986).
10. Katzen, A. L., Kornberg, T. B. & Bishop, J. M. *Cell* **41**, 449-456 (1985).

GUIDE TO AUTHORS

Please follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers for whom English is a second language and those who work in other fields. Please write clearly and simply, avoiding unnecessary technical terminology. *Nature's* staff will edit manuscripts to those ends if necessary. Contributors should check their proofs carefully.

Because of the competition for space, many of the papers submitted for publication cannot be accepted. For this reason, and because brevity is a great assistance to readers, papers should be as brief as is consistent with intelligibility. Please note that one printed page of *Nature*, without diagrams or other interruptions of the text, has fewer than 1,300 words.

CATEGORIES OF PAPER

Review Articles survey recent developments in a field. Most are commissioned, but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance but must not exceed six pages of *Nature*.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text or more than six display items (figures and tables) and should not occupy more than five pages of *Nature*.

Articles should be accompanied by a heading of 50–80 words written to advertise their contents in general terms, to which editors will pay particular attention. A heading (printed in italic type) is not an abstract and should not usually contain numbers or measurements. The study should be introduced in more detail in the first two or three paragraphs, which should also briefly summarize its results and implications.

Articles may contain a few subheadings of two or three words. The meaning of the text should not depend on the subheadings, whose function is to break up the text and to point to what follows. There should be fewer than 50 references.

Letters to *Nature* are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should not have more than 1,000 words of text or more than four display items and should not occupy more than two pages of *Nature*. The first paragraph should describe, in not more than 150 words, the origins and chief conclusions of the study. Letters should not have subheadings or more than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned, most are unsolicited. They are normally between one and four pages of *Nature* in length.

News and Views articles are intended to inform non-specialist readers about a recently published advance. Suggestions should be made to the News and Views Editor. Illustrations are welcome. Proposals for meeting reports should be agreed in advance.

Scientific Correspondence is for the discussion of scientific matters, including contributions published in *Nature*. Priority is given to contributions of less than 500 words and five references. Figures are welcome.

Submission. All manuscripts may be submitted to either London or Washington. They should not be addressed to editors by name. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax (VAT).

GENERAL

All contributions submitted for publication in *Nature* should conform with these rules.

Manuscripts should be typed, double-spaced, with a good-quality printer, on one side of the paper only. Four copies are required, each accompanied by lettered artwork. Four copies of half-tones should be included. Reference lists, figure legends and so on should be on separate sheets, all of which should be double-spaced and numbered. Copies of relevant manuscripts in press or submitted for publication elsewhere should be included, clearly marked as such. Revised and resubmitted manuscripts should also be clearly marked as such and labelled with their reference numbers.

Titles should say what the paper is about with the minimum of technical terminology and in fewer than 80 characters. Authors should avoid active verbs, numerical values, abbreviations and punctuation and should include one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name, figure number and, when known, manuscript reference number. Original artwork should be unlettered. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are permitted only if the parts are closely related, either experimentally or logically. Suggestions for cover illustrations, with captions, are welcome. Original artwork (and one copy of the manuscript) will be returned when a manuscript cannot be published.

Colour Artwork. A charge of £500 a page is made for 4-colour figures. Inability to meet these costs will not prevent the publication of essential colour figures if the circumstances are explained.

Figure legends must not exceed 300 words and ideally should be much shorter, and should use telegraphic form wherever possible. The figure should be described first, then, briefly, the method. Reference to a method described elsewhere is preferable to a full description. Methods should not be described in detail in the text.

References should be numbered sequentially as they appear in the text, followed by those in tables and finally those in the figure legends. Reference numbers apply only to papers published or in the press, and a different number should be given to each paper cited. All other forms of reference (including unrefereed abstracts) should be included in the text as personal communication, manuscript in preparation or preprint (with number and institution where appropriate). Text should not be included in the reference list. References should be abbreviated according to the *World List of Scientific Periodicals* (fourth edition, Butterworth, London, 1963–65). The first and last page numbers should be cited. Reference to books should clearly indicate the publisher and the date and place of publication.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided as much as possible and, when used, defined. Editors will shorten words if necessary. In case of doubt, SI units should be used.

Footnotes should not be used except for changed addresses or to identify the corresponding author if different from the first-named.

Acknowledgements should be brief. Grant numbers and contribution numbers are not allowed.

Receptor-affinity chromatography

P. Bailon and D.V. Weber

Based upon the specificity and reversibility of receptor-ligand interactions, receptor-affinity chromatography is one of the newest affinity methods for the purification of biomolecules.

AFFINITY chromatography, which utilizes the ability of macromolecules in solution to bind specifically and reversibly to immobilized ligands^{1,2}, became a fully-fledged technology in the 1970s. Before this time, biomolecules such as enzymes, hormones, receptors and antibodies were purified by traditional chromatographic methods. Affinity chromatography is a facile purification method and its applications are virtually limitless. Immunoaffinity chromatography, based on antigen-antibody interactions, became very popular after the discovery of monoclonal antibody-producing hybrid cell lines by Kohler and Milstein³ in 1975.

Previously, we have reported the development of a novel affinity purification method termed "receptor-affinity" chromatography⁴. Receptor-affinity chromatography (RAC) exploits the specificity and reversibility of the interactions between a matrix-bound receptor and its soluble protein ligand.

The unique properties of receptor-ligand interactions impart several useful characteristics to RAC. The receptor-affinity adsorbents selectively bind with high avidity only the fully active biomolecule in its native conformation, and the receptor-ligand complex formed is easily dissociated with mild reagents, allowing high recoveries of the bound biologically active protein. These practical characteristics make RAC an ideal tool for the downstream purification of recombinant proteins from microbial and mammalian sources.

Purification of rIL-2

To date, recombinant human interleukin-2 (rIL-2), its various analogues, a recombinant murine homologue (unpublished data) and an interleukin-2-*Pseudomonas* exotoxin fusion protein (IL-2-PE40) have been purified employing this technology^{5,6}. A recombinantly produced soluble form of the low-affinity p55 subunit of the human interleukin-2 receptor^{7,8} was immobilized to the commercially available, activated NuGel P-AF NHS support (Separation Industries, Metuchen, New Jersey), as described⁵. The NuGel support was chosen for its highly porous and non-compressible nature, and its high flow properties — important factors to consider when scaling-up purification processes.

The receptor-affinity beads were packed into a chromatographic column. After equilibration with phosphate-

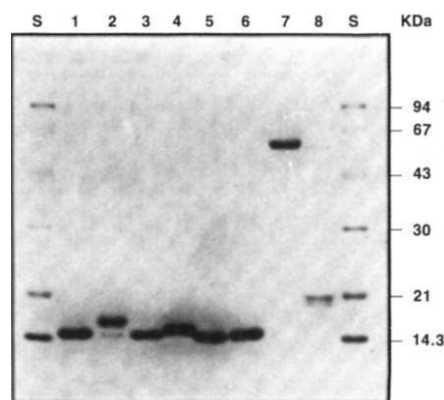


Fig. 1 Receptor-affinity purified recombinant interleukin-2 (rIL-2) and its analogues. Track S: standard molecular weight marker proteins. Tracks 1-8: microbial rIL-2, mammalian glycosylated rIL-2, yeast rIL-2, rIL-2 mutein Lys(20), rIL-2 mutein Des-Ala(1)-Ser(125), Seleno-Met-rIL-2, IL-2-PE40 and murine rIL-2, respectively.

buffered saline (PBS) buffer (pH 7.4) the crude rIL-2 preparations from microbial, mammalian and yeast origin were passed through the column. The receptor column was operated at or above its saturating binding capacity so that it would preferentially bind the high-affinity, fully-active soluble monomeric form of rIL-2.

Immediately after adsorption, the column was washed extensively with PBS buffer to remove unadsorbed and non-specifically adsorbed materials from the beads. The specifically-adsorbed rIL-2 was then eluted with 0.2 M acetic acid containing 0.2 M sodium chloride. The rIL-2 was concentrated and subjected to gel filtration in Sephadex G-50 (Pharmacia LKB, Bromma, Sweden) using 50 mM sodium acetate containing 5 mg mannitol per millilitre and 0.2 M sodium chloride (pH 3.5). During this step, no significant increase in purity is usually achieved, but this is a convenient way to exchange the product into the final storage buffer.

rIL-2 analogues

Employing recombinant DNA technology, three human rIL-2 analogues (rIL-2 Lys(20) mutein, rIL-2 Des-Ala(1)-Ser(125) mutein and Seleno-Met-rIL-2), as well as a murine homologue and an interleukin-2-*Pseudomonas* exotoxin fusion protein (IL-2-PE40), were engineered, expressed in *Escherichia coli*, and purified by receptor-affinity chromatography. In the rIL-2 Lys(20) mutein, the

aspartic acid at position 20 in the human IL-2 is substituted with a lysine residue. This single amino acid substitution results in a thousand-fold reduction in bioactivity, with no reduction in binding to the p55 subunit of the receptor. This has been interpreted as a demonstration that the aspartic acid at position 20 is critical for the interaction of IL-2 with the p75 receptor subunit of the high-affinity complex⁹.

Similarly, in the rIL-2 Des-Ala(1)-Ser(125) mutein, alanine at position 1 in the human IL-2 is deleted and cysteine at position 125 is replaced with a serine residue. The seleno-methionine IL-2 was used in the X-ray crystallographic structure determination of rIL-2 (B. Graves and M. Hatada, personal communication). The selenium atom in the rIL-2 causes an anomalous scattering of the diffracted beam, which is very useful in X-ray crystallographic studies at multiple wavelengths. The IL-2-PE40 is a chimaeric protein in which the cell recognition domain of the *Pseudomonas* exotoxin was replaced with IL-2¹⁰. IL-2-PE40 is a cytotoxic agent for cells bearing IL-2 receptors. Consequently, IL-2-PE40 may be a useful therapeutic agent in the treatment of diseases in which IL-2 receptors play a negative role, such as in organ transplantation rejection and autoimmune diseases.

The SDS-PAGE profiles of the RAC-purified rIL-2s are shown in Fig. 1. The high purity of the receptor-affinity purified materials is indicated by the absence of contaminating proteins. RAC has provided sufficient quantities of highly purified and biologically active rIL-2s and analogues for preclinical testing and structure-function studies.

Comparison

The gel filtration profiles of receptor- and immunoaffinity-purified rIL-2 are compared in Fig. 2. The RAC-purified rIL-2 (Fig. 2a) contains essentially the soluble monomeric form with little or no oligomeric or aggregated forms (high-molecular-weight peak). In contrast, the immunoaffinity-purified material (Fig. 2b) contains significant amounts of oligomers in addition to the desired monomeric form. The results clearly demonstrate that the receptor column preferentially binds the soluble monomeric form of rIL-2 from a heterogeneous population of other IL-2 molecules, whereas the immunoaffinity column binds various molecular forms of rIL-2 of varying

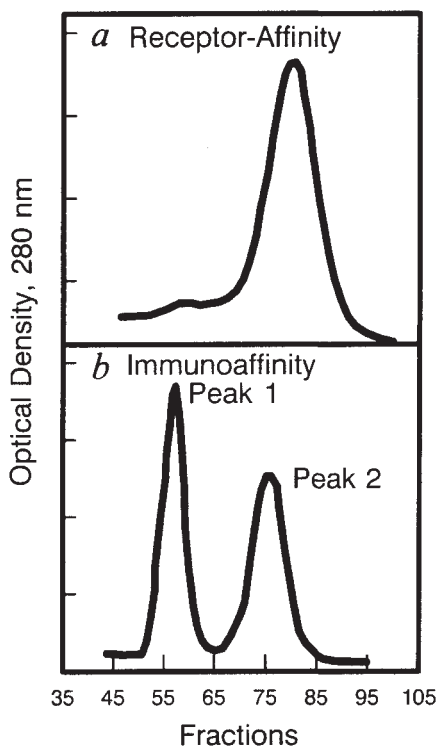


Fig. 2 Gel filtration profiles of receptor- and immunoaffinity-purified rIL-2^s. A 2.6 × 35 cm column is packed with superfine Sephadex G-50. The mobile phase used was 50 mM sodium acetate containing 0.2 M sodium chloride, 5 mg ml⁻¹ mannitol, pH 3.5. Eleven mg protein in 3 ml buffer was applied to the column at a flow rate of 0.5 ml min⁻¹ and 10-minute fractions were collected.

degrees of biological activity and renaturation state.

The widespread commercial use of RAC depends upon considerable reductions in the production costs of soluble receptors. Good Manufacturing Practices criteria for the use of the receptors also have yet to be established.

Biotechnology is on the verge of producing soluble receptors of other biomolecules such as interleukin-1¹, gamma-interferon and tumour necrosis factor. We believe that in the near future receptor-affinity chromatography will become an established method for the purification of high-value recombinant proteins. □

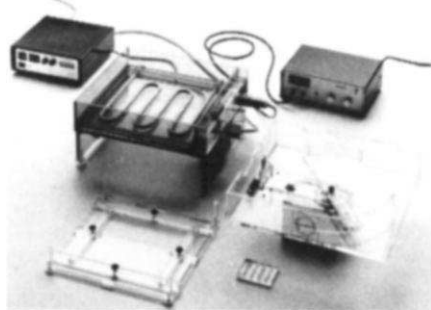
Pascal Bailon and David V. Weber are at the Protein Biochemistry Department, Hoffmann-La Roche, Roche Research Center, Nutley, New Jersey 07110, USA. For more information, fill in reader service number 100.

1. Cuatrecasas, P., Wilchek, M. & Anfinsen, C.B. *Proc. natn. Acad. Sci. U.S.A.* **61**, 636-643 (1968).
2. Cuatrecasas, P. & Anfinsen, C.B. *A. Rev. Biochem.* **40**, 259-278 (1971).
3. Lowe, C.R. *Int. J. Biochem.* **8**, 177-181 (1977).
4. Kohler, G. & Milstein, C. *Nature* **256**, 975-9754 (1975).
5. Bailon, P. *et al. Bio/Technology* **5**, 1195-1198 (1987).
6. Bailon, P. *et al. Bio/Technology*, (in the press).
7. Hakimi, J. *et al. J. biol. Chem.* **262**, 17336-17341 (1987).
8. Weber, D.V., Keeney, R.F., Familletti, P.C. & Bailon, P. *J. chrom. biomed. Applic.* **431**, 55-63 (1988).
9. Collins, L. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).
10. Lorberbaum-Galski, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 1922-1926 (1988).
11. Sims, J.E. *et al. Science* **241**, 585 (1988).

Genes and things

A handful of reagents and hardware for 'genetically engineering' cells are offered this week, including competent cells, a labelled nucleotide and a restriction mapping kit.

THE German company Biometra has come up with a new slant on **pulsed field gel electrophoresis** techniques (*Reader Service No. 101*). Biometra's Rotaphor electrophoresis apparatus has rotating



Biometra's Rotaphor electrophoresis apparatus.

electrodes which are curved to generate a homogeneous field between them. A stepping motor, controlled by a microprocessor-based unit, alters the orientation of the electric field at arbitrary predetermined angles at set intervals of between 5 and 9,999 seconds. The company says DNA molecules of up to 10⁶ base pairs can be separated on a 20 × 20 cm agarose gel in 24 hours, without lane broadening and fanning.

ICN Biomedicals offers ³⁵S-labelled L-cysteine for *in vitro* translation using rabbit reticulocyte lysate, wheat germ lysate and RNA-dependent reticulocyte lysate systems (*Reader Service No. 102*). The reagent can also be used to label mammalian cells in culture and proteins *in vivo*. ICN Biomedicals sells the compound as a 1 mCi aqueous solution containing 10 mM 2-mercaptoethanol in a multipurpose vial, for the introductory price of \$95 (US), valid until the end of the year.

Clontech has introduced a line of organic reagents to automate the **modification of synthetic oligonucleotides** in preparation for non-isotopic labelling (*Reader Service No. 103*). The line includes \$75-90 (US) C₆ AminoModifiers, \$85 (US) AminoModifier II and \$75 (US) C₆ ThiolModifier for the attachment of a primary, aliphatic amine or sulphhydryl group to the 5' terminus or an internal position of a DNA sequence. The products come as CE-phosphoramidites

These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.

which plug in to the fifth port of a DNA synthesizer, along with the standard G, A, T and C phosphoramidites.

BRL says its new MAX Efficiency DH5α5F'IQ **competent cells** deliver the highest transformation efficiency available for M13/phagemid cloning (*Reader Service No. 104*). The competent cells are supplied with lawn cells and yield > 1 × 10⁸ plaque-forming units per microgram of monomer M13mp19 RF DNA or > 3 × 10⁸ transformants per microgram pBR322 DNA, says the company. The strain contains a kanamycin resistance marker to prevent loss of the F' episome, and the cells overproduce the *lac* repressor protein, making them useful for cloning into expression vectors.

Promega has added the *Sfi*I high-resolution **mapping system** to its line of molecular biology products (*Reader Service No. 105*). Promega's LambdaGEM-11 genomic cloning vector possesses *Sfi*I restriction sites flanking bacteriophage T7/SP6 RNA polymerase promoters and a multiple cloning region. The *Sfi*I sites allow most inserts to be excised as a single fragment, because the 8-base restriction site occurs infrequently in genomic DNA. Promega has designed the *Sfi*I sites in an asymmetric fashion, so that radiolabelled linkers complementary to either terminus can be separately ligated to the *Sfi*I excised genomic DNA, and a restriction map can be determined by partial digestion with a frequent-cutting restriction endonuclease. The company says the mapping resolution of the method is much greater than conventional *cos* site oligo-labelling because only the ends of the insert are labelled, instead of the 20 kilobase and 9 kilobase arms of the vector. The complete system performs 200 reactions, and sells for \$265 (US). □

ADVERTISEMENT

Medwire Div.
stocks a wide variety of
TEFLON COATED*

Platinum alloys, stainless steel, silver, multistrand and single strand available for prompt shipment. Write for descriptive brochure and price lists



Medwire Div.
Affiliate of Sigmund Cohn Corp
123 So. Columbus Ave.
Mt. Vernon, N.Y. 10553
914-664-5300
*Reg. T.M. of E.I. DuPont
Reader Service No.49